

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2003, 16:56:01 ; Search time 1223 Seconds
(without alignments)
11515.101 Million cell updates/sec

Title: US-09-996-956-1
Perfect score: 5217
Sequence: 1 gaactttaaataatccctc.....taaaaaaaaaaaaaaaaaaaaaa 5217

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03.*
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2525.4	48.4	11801	21 AAC61388	An androgen-related
2	1531.6	29.4	1579	21 AAC61382	CDNA sequence of a
3	1456.8	27.9	1603	21 AAC61381	CDNA sequence of a
C 4	1308	25.1	1804	23 ABV23954	Human prostate exp
C 5	1308	25.1	1804	23 ABV24032	Human prostate exp
C 6	1308	25.1	1804	23 ABV29837	Human prostate exp
C 7	1308	25.1	1804	23 ABV29916	Human prostate exp
8	891.4	17.1	1025	21 AAC61384	Probe for androgen

C	9	462	8.9	249999	25	ABZ80229	Human tramdorin ge
	10	454	8.7	2612	22	AAH17215	Human cDNA sequenc
	11	438.4	8.4	843	24	ABQ89916	Human prostate exp
C	12	421.2	8.1	33206	24	ABN95882	Gene #2380 used to
	13	417.6	8.0	1761	21	AAV74309	Human secreted pro
C	14	404.6	7.8	80240	20	AAV83940	NC-contig derived
C	15	404.6	7.8	80595	20	AAV83939	HC-contig derived
	16	390.6	7.5	27666	25	ABZ74190	Secreted protein g
	17	390.6	7.5	27666	25	ABZ67756	Human secreted pro
C	18	390.6	7.5	32212	22	AAH06082	Human secreted pro
C	19	390.6	7.5	32212	23	ABL98647	Human reproductive
	20	390.6	7.5	41206	25	ABZ74191	Human testicular a
	21	390.6	7.5	41206	25	ABZ67757	Secreted protein g
	22	390.4	7.5	848	24	ABQ89971	Human secreted pro
	23	386.6	7.4	820	24	ABQ89731	Human prostate exp
C	24	386.2	7.4	6138	22	AAH05378	Human reproductive
C	25	386.2	7.4	6138	23	ABL98624	Human testicular a
	26	380.4	7.3	790	24	ABQ89552	Human prostate exp
	27	377.8	7.2	802	24	ABQ89827	Human prostate exp
	28	374.8	7.2	237961	24	ABQ80552	Human prostate gene
	29	363.6	7.0	1691080	24	ABX08336	Human phosphodiester
	30	359.8	6.9	599	22	ABA62069	Human foetal liver
	31	359.8	6.9	599	22	ABA29520	Probe #7986 for ge
	32	359.8	6.9	599	22	AAK10387	Human brain expres
	33	359.8	6.9	599	22	AAK36288	Human bone marrow
	34	359.8	6.9	599	22	AAI17277	Probe #7210 for ge
	35	359.8	6.9	599	22	AAI42011	Probe #10697 used
	36	359.8	6.9	599	23	ABS35968	Human liver single
	37	359.8	6.9	599	24	ABS10371	Human genome-deriv
C	38	357.8	6.9	438	23	ABV36282	Human prostate exp
C	39	357.8	6.9	438	23	ABV45318	Human prostate exp
C	40	357.4	6.9	409	23	ABV15490	Human prostate exp
C	41	351	6.7	1650	24	ABL49826	Human Ras GTP enzy
C	42	350.8	6.7	1980	22	AAK94596	Human full-length
C	43	350	6.7	28818	22	AAH35901	Human musculoskele
	44	350	6.7	28818	25	ABX58889	CDNA encoding nove
	45	338.8	6.5	154465	24	AAZ8763	Human AKAP allelic

ALIGNMENTS

RESULT 1

AAC61388

ID AAC61388 standard; DNA; 11801 BP.

XX AAC61388;

AC AAC61388;

XX 19-FEB-2001 (first entry)

DT An androgen-related, prostate-specific gene PCGEM1.

DE Androgen-regulated gene; prostate specific gene; PCGEM1; prostate cancer;

XX prostate cancer gene expression marker 1; prostate related disease;

KW benign prostatic hyperplasia; ss.

XX Homo sapiens.

OS WO200058470-A1.

PN 05-OCT-2000.

PD 24-MAR-2000; 2000WO-US07906.

PF 26-MAR-1999; 99US-0126469.

PR (SRK//) SRIKANTAN V.

XX (ZOUZ//) ZOU Z.

PA (MOUL//) MOUL J W.

XX (SRIV//) SRIVASTAVA S.

PI Srikantan V, Zou Z, Moul JW, Srivastava S;

DR WPI; 2000-664926/64.

XX Novel androgen-regulated prostate specific gene, prostate cancer gene

PT expression marker, useful for detecting, diagnosing, preventing, and

PT treating prostate cancer and other prostate related diseases -

XX PS

XX Claim 1; Fig 14; 75pp; English.

XX The present sequence is the genomic DNA sequence of androgen-regulated,

CC prostate specific gene PCGEM1 (prostate cancer gene expression marker 1).

CC The PCGEM1 gene is over-expressed in prostate cancer. The PCGEM1

CC polynucleotide is useful for detecting prostate cancer in a patient.

CC The PCGEM1 promoter may be linked to cytotoxic gene, and be used

CC for selectively killing prostate cancer cells. The PCGEM1 polynucleotide

CC is also useful as marker of prostate cancer and other prostate related

CC diseases, as targets for therapeutic intervention in prostate cancer

CC and other prostate related diseases, in detection, diagnosis, prognosis,

CC prevention, and treatment of prostate cancer (e.g. prostatic

CC intraepithelial neoplasia (PIN), adenocarcinomas, nodular hyperplasia,

CC and large duct carcinomas) and prostate related diseases (e.g. benign

CC prostatic hyperplasia).

XX PS

SQ Sequence 11801 BP; 3740 A; 2033 C; 1968 G; 4059 T; 1 other;

Query Match 48.4%; Score 2525.4; DB 21; Length 11801;

Best Local Similarity 97.8%; Pred. No. 0;

Matches 2754; Conservative 0; Mismatches 36; Indels 27; Gaps 18;

QY 423 AAGTAACGGGATACAGCAGGCTGGTGGCATTTGACATTCATGATATCAGCCAAAGTG 482

DB 8981 AATATTTTCACCCACAGGCTGGTGGCATTTGACATTCATGATATCAGCCAAAGTG 9040

QY 483 GAACATAAAACAGCTCCTGGAAGGACTATGACATCATCAGGTTGGAGTCTCCAGGGA 542

DB 9041 GAACATAAAACAGCTCCTGGAAGGACTATGACATCATCAGGTTGGAGTCTCCAGGGA 9100

QY 543 CAGCGGACCTTTGGAAAGGACTAGAAAGTGTGAATCTATTAGTCTTCGATATGAAT 602

DB 9101 CAGCGGACCTTTGGAAAGGACTAGAAAGTGTGAATCTATTAGTCTTCGATATGAAT 9160

QY 603 TCCTCTCTCTCTG-AAAAGCATTTCAATTTTACAGACACAGGCTACTCTAGGGCAGC 661

DB 9161 TCCTCTCTCTCTG-AAAAGCATTTCAATTTTACAGACACAGGCTACTCTAGGGCAGC 9220

QY 662 AAAAAGTGGCAACAGGCAAGCAGAGGAAAAGAGATCATCAGGCAATTCAGAGTGCACGTG 721

DB 9221 AAAAAGTGGCAACAGGCAAGCAGAGGAAAAGAGATCATCAGGCAATTCAGAGTGCACGTG 9280

QY 722 TCCTTTTCATATATTTCTCAATGCCGTATGTTGGTTTATTTTGGCCAAAGCATCAACATC 781

DB 9281 TCCTTTTCATATATTTCTCAATGCCGTATGTTGGTTTATTTTGGCCAAAGCATCAACATC 9340

QY 782 TGCTCAAGAAAAAATCTGGAGAAAACAAAGTGCTTTGGCAATGTTATGTTCTTTT 841

DB 9341 TGCTCAAGAAAAAATCTGGAGAAAACAAAGTGCTTTGGCAATGTTATGTTCTTTT 9400

QY 842 TTGACAAGCCCTCAGATTTCTGAGGGGAATTCACATAAATGGGATCAGGTCATTCATTTA 901

DB 9401 TTGACAAGCCCTCAGATTTCTGAGGGGAATTCACATAAATGGGATCAGGTCATTCATTTA 9460

QY 902 CGTTGTGTGCAATATGATTTAAAGATACAACTTTTGCAGAGAGCATGCTTTCCCTAAGGG 961

DB 9461 CGTTGTGTGCAATATGATTTAAAGATACAACTTTTGCAGAGAGCATGCTTTCCCTAAGGG 9520

QY 962 TAGGCACGTGGAGGACTAAGGTTAAAGCATTTCTTCAAG-ATCAGTTAATC-AGAAAGGT 1019

DB 9521 TAGGCACGTGGAGGACTAAGGTTAAAGCATTTCTTCAAGATCAGTTAAATCAAGAAAGGT 9580

QY 1020 GCTCTTTTGCATTTCTGAAATGCCCTTGTGCAAAATATTGGTTATATTGATTAATTTTACAC 1079

DB 9581 GCTCTTTTGCATTTCTGAAATGCCCTTGTGCAAAATATTGGTTATATTGATTAATTTTACAC 9640

QY 1080 TTAATGGAACAACACCTTTAACTTACAGATGAACAAACCCCAAAAGCAAAAATCAAAAG 1139

DB 9641 TTAATGGAACAACACCTTTAACTTACAGATGAACAAACCCACAAAAGCAAAAATCAAAAG 9700

QY 1140 CCCTACCTATGATTTTCATATTTCTGTGTAAGTAAAGGATTCCTGCTGCTTTTG 1199

DB 9701 CCCTACCTATGATTTTCATATTTCTGTGTAAGTAAAGGATTCCTGCTGCTTTTG 9760

QY 1200 GGCAATAATGATAATGGAATATTTCCAGGTATTTTAAATAGGGCCCATCTACAAT 1259

DB 9761 GGCAATAATGATAATGGAATATTTCCAGGTATTTTAAATAGGGCCCATCTACAAT 9820

QY 1260 TCTTAGCAATACCTTTGGATATTTCAAAATTCAGCTGGACATTTGCTAAATTTTAT 1319

DB 9821 TCTTAGCAATACCTTTGGATATTTCAAAATTCAGCTGGACATTTGCTAAATTTTAT 9880

QY 1320 ATACATCTTTGCTAGATTTTCAAAATTTAAGTATTTAGTAAATTTAGTAAATTTAGT 1379

DB 9881 ATACATCTTTGCTAGATTTTCAAAATTTAAGTATTTAGTAAATTTAGTAAATTTAGT 9940

QY 1380 ATCAATTCAAAACATTTACTTTTCCATAATTTAGACTATGAAGTTCATAAATTCACAAA 1439

DB 9941 ATCAATTCAAAACATTTACTTTTCCATAATTTAGACTATGAAGTTCATAAATTCACAAA 10000

QY 1440 TATATCTACACATACAATTTATAGATTTGTTTTCATTTATATATATATATATATATATAT 1499

DB 10001 TATATCTACACATACAATTTATAGATTTGTTTTCATTTATATATATATATATATATATAT 10060

QY 1500 TGCTTTTGTGATTTTGTAGAAAACCTGAGAGTTTTTAATTCATCAATTAATTTAC-TTGATCAAAA 1558

DB 10061 TGCTTTTGTGATTTTGTAGAAAACCTGAGAGTTTTTAATTCATCAATTAATTTACGTTGATCAAAA 10120

QY 1559 AATGTGGGAACAACTCAGCATTAATTTGTATGTGATTTTATATATATATATATATATATAT 1618

DB 10121 AATGTGGGAACAACTCAGCATTAATTTGTATGTGATTTTATATATATATATATATATAT 10180

QY 1619 TAAGCTTTGGTGGCTTGAAGTCTTTTGTACTTGTAGTCCCATGTTTAAATTTACTACTTATA 1678

DB 10181 TAAGCTTTGGTGGCTTGAAGTCTTTTGTACTTGTAGTCCCATGTTTAAATTTACTACTTATA 10240

QY 1679 TCTTAAGCATTTATGTTTTCATTAATTTTACATGATGCTAAATTTATGGAATTTATATATAT 1738

DB 10241 TCTTAAGCATTTATGTTTTCATTAATTTTACATGATGCTAAATTTATGGAATTTATATATAT 10300

QY 1739 AATATTAAAGATTTTCAAAATAGATATGTAATTTGTTTCA-CATACATAGAAATGAAAGT 1797

DB 10301 AATATTAAAGATTTTCAAAATAGATATGTAATTTGTTTCAACATACATAGAAATGAAAGT 10360

QY 1798 TCATTTGCTAAAGCAAGATGCTGGTGAAGAGTCTTTTGTATGTAAGATCAGTACTAGATT 1857

DB 10361 TCATTTGCTAAAGCAAGATGCTGGTGAAGAGTCTTTTGTATGTAAGATCAGTACTAGATT 10420

QY 1858 AGTAGAGGCAAGACTTCTAGTCCCTAATCTACCTTTAATAGCCATGCTGCTAGCTGTAA 1917

DB 10421 AGTAGAGGCAAGACTTCTAGTCCCTAATCTACCTTTAATAGCCATGCTGCTAGCTGTAA 10480

QY 1918 GTCAAGTGAACCCATCTCATTTCTCTCATCTTTTTCATCTCTCTCTCTCTCTCTCTCTCTCT 1977

DB 10481 GTCAAGTGAACCCATCTCATTTCTCTCATCTTTTTCATCTCTCTCTCTCTCTCTCTCTCTCT 10540

QY 1978 TAAGCTC-TTCATTTTCTTTTCTTTTGTAGATAGATTTTGTCTCTCTCTCTCTCTCTCTCTCTCT 2036

DB 10541 TAAGCTCCTTTTCTTTTCTTTTGTAGATAGATTTTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 10600

QY 2037 AGTGCATGSCAGCATCTCAGCTCAGCTGCAACCTCTGCTTCTCTCTCTCTCTCTCTCTCTCTCT 2096

DB 10601 AGTGCATGSCAGCATCTCAGCTCAGCTGCAACCTCTGCTTCTCTCTCTCTCTCTCTCTCTCTCT 10660

QY 2097 -CCTGCTTACGCTCCCAAGT---AGCGGGATTTACAGGTGCCGCCACCACCATCTGG--- 2150

DB 10661 CCCTGCTTACGCTCCCAAGTGGCGGGATTTACAGGTGCCGCCACCACCATCTGGGCC 10720

QY 2151 CTAAATTTTGTATTTTTCACCATGTTGGCCAGGCTGCTCGAACCCCTACCTACCTAGTGA 2210

Qy 456 TGACACTTCATGATATCAGCCAAAGTGGAACTAAAAACAGCTCCTCGGAAGAGGACTATGA 515
Db 255 TGACACTTCATGATATCAGCCAAAGTGGAACTAAAAACAGCTCCTCGGAAGAGGACTATGA 314
Qy 516 CATCATCAGTTGGAGTCTCCAGGACAGCGGACCCCTTTGGAAAAGGACTAGAAAGTGT 575
Db 315 CATCATCAGTTGGAGTCTCCAGGACAGCGGACCCCTTTGGAAAAGGACTAGAAAGTGT 374
Qy 576 GAAATCTATTAGTCTTCGATATGAAATCTCTGCTCTGTAAAGACATTTTCATATTTTACA 635
Db 375 GAAATCTATTAGTCTTCGATATGAAATCTCTGCTCTGTAAAGACATTTTCATATTTTACA 434
Qy 636 AGACACAGGCTACTCTAGGCGCAGCAAAAAGTGGCAACAGCGCAAGCAGAGGGAAGAG 695
Db 435 AGACACAGGCTACTCTAGGCGCAGCAAAAAGTGGCAACAGCGCAAGCAGAGGGAAGAG 494
Qy 696 ATCATGAGGATTTTCAGATGCACTGTCTTTTCATATATTTCTCAATGCCGTATGTTGG 755
Db 495 ATCATGAGGATTTTCAGATGCACTGTCTTTTCATATATTTCTCAATGCCGTATGTTGG 554
Qy 756 TTTTATTTTGGCCAAAGCATACATCTGCTCAAGAAAAAATCTGGAGAAAAACAAGG 815
Db 555 TTTTATTTTGGCCAAAGCATACATCTGCTCAAGAAAAAATCTGGAGAAAAACAAGG 614
Qy 816 TGCCTTTGCCAATGTTATGTTTCTTTTGACAAGCCCTGAGATTTCTGAGGGGAATTCAC 875
Db 615 TGCCTTTGCCAATGTTATGTTTCTTTTGACAAGCCCTGAGATTTCTGAGGGGAATTCAC 674
Qy 876 ATAAATGGGATCAGGTCATTCATTAGCTTGTGCAAAATATGATTTAAAGATACACCT 935
Db 675 ATAAATGGGATCAGGTCATTCATTAGCTTGTGCAAAATATGATTTAAAGATACACCT 734
Qy 936 TTGCAGAGAGCATGCTTTCTTCAAGGGTAGGCAGCTGAGGACTAAGGTTAAAGCATTCCT 995
Db 735 TTGCAGAGAGCATGCTTTCTTCAAGGGTAGGCAGCTGAGGACTAAGGTTAAAGCATTCCT 794
Qy 996 CAAGATCAGTTAATCAAGAAGTGCCTTTTGGCAATCTGAAATGCCCTTGTGCAAAAT 1055
Db 795 CAAGATCAGTTAATCAAGAAGTGCCTTTTGGCAATCTGAAATGCCCTTGTGCAAAAT 854
Qy 1056 TGCTTATGATTAAATTTACACTTAATGGAACACCTTTTACCTTACAGATCAACAAA 1115
Db 855 TGGTTATGATTAAATTTACACTTAATGGAACACCTTTTACCTTACAGATCAACAAA 914
Qy 1116 -CCACAAAAGCAAAAATCAAAAGCCCTACCTATGATTTTCATATTTTCTGTGTAACCTG 1174
Db 915 CCCACAAAAGCAAAAATCAAAAGCCCTACCTATGATTTTCATATTTTCTGTGTAACCTG 974
Qy 1175 ATTAAGGATTCCTGCTTGTCTTTTGGGCATAAATGATAATGGAATTTCCAGGATTTGT 1234
Db 975 ATTAAGGATTCCTGCTTGTCTTTTGGGCATAAATGATAATGGAATTTCCAGGATTTGT 1034
Qy 1235 TTAATATGAGGCCCTCACTACAATCTTACCAATCTTGGATAATTTCTAAATTCACG 1294
Db 1035 TTAATATGAGGCCCTCACTACAATCTTACCAATCTTGGATAATTTCTAAATTCACG 1094
Qy 1295 TGGACATGTCATATTTTATATACATCTTTGCTAGAAATTTCAAAATTTAAAGTATG 1354
Db 1095 TGGACATGTCATATTTTATATACATCTTTGCTAGAAATTTCAAAATTTAAAGTATG 1154
Qy 1355 TGAATTTAGTTAATAGCTGTGCTGATCAATCAAAACAAATTTCTTTCCFAAATTTTGA 1414
Db 1155 TGAATTTAGTTAATAGCTGTGCTGATCAATCAAAACAAATTTCTTTCCFAAATTTTGA 1214
Qy 1415 CTATGAGGTCATAAATTCACAAATATATCTACACATCAAAATATAGATTTTTCAT 1474
Db 1215 CTATGAGGTCATAAATTCACAAATATATCTACACATCAAAATATAGATTTTTCAT 1274
Qy 1475 TATAATGCTCTCATCTTAACAGAAATGCTCTTTGTTGTTGTTTTAGAAAACTGAGATTT 1534
Db 1275 TATAATGCTCTCATCTTAACAGAAATGCTCTTTGTTGTTGTTTTAGAAAACTGAGATTT 1334
Qy 1535 TAATTCATATTAATCTTGATCAAAAAAATTTGTGGGAACAATCCAGCATTAATTTGATGTGAT 1594

Db 1335 TAATTCATAATTAATCTTGATCAAAAAATTTGGGAACAATCCAGCATTAATTTGATGTGAT 1394
Qy 1395 TGTATTTATGTACATAAGGAGTCTTAAGCTTTGGTCCCTTGAAGTCTTTTGTACTTAGTCC 1654
Db 1395 TGTATTTATGTACATAAGGAGTCTTAAGCTTTGGTCCCTTGAAGTCTTTTGTACTTAGTCC 1454
Qy 1655 CATGTTTAAAAATTAATCTACTTTATATCTAAAGCATTTATGTTTTCATTTCAATTTACATG 1714
Db 1455 CATGTTTAAAAATTAATCTACTTTATATCTAAAGCATTTATGTTTTCATTTCAATTTACATG 1514
Qy 1715 ATGCTAATTAATGCGCAATTTATTAACAATATTAAGATTTTCGAAATAGATAA 1764
Db 1515 ATGCTAATTAATGCGCAATTTATTAACAATATTAAGATTTTCGAAATAGATAA 1564

RESULT 3
AAC61381
ID AAC61381 standard; cDNA; 1603 BP.
XX
AC AAC61381;
XX
AC AAC61381;
DT 19-FEB-2001 (first entry)
XX
cDNA sequence of androgen-related, prostate-specific gene PCGEM1.
DE Androgen-regulated gene; prostate specific gene; PCGEM1; prostate cancer;
XX prostate cancer gene expression marker 1; prostate related disease;
KW benign prostatic hyperplasia; ss.
XX
OS Homo sapiens.
XX
PN WO200058470-A1.
XX
PD 05-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US07906.
XX
PR 26-MAR-1999; 99US-0126469.
XX
PA (SRIK/) SRIKANTAN V.
PA (ZOUZ/) ZOU Z.
PA (MOUL/) MOUL J W.
PA (SRIV/) SRIVASTAVA S.
XX
PI Srikantan V, Zou Z, Moul JW, Srivastava S;
XX
DR WPI: 2000-664926/64.
XX
PT Novel androgen-regulated prostate specific gene, prostate cancer gene
expression marker, useful for detecting, diagnosing, preventing, and
PT treating prostate cancer and other prostate related diseases
XX
PS Claim 1; Fig 8; 75pp; English.
XX
CC The present sequence represents a cDNA fragment of an androgen-regulated,
CC prostate specific gene PCGEM1 (prostate cancer gene expression marker 1).
CC The PCGEM1 gene is over-expressed in prostate cancer. The PCGEM1
CC polynucleotide is useful for detecting prostate cancer in a patient.
CC The PCGEM1 promoter may be linked to cytotoxic gene, and be used
CC for selectively killing prostate cancer cells. The PCGEM1 polynucleotide
CC is also useful as marker of prostate cancer and other prostate related
CC diseases, as targets for therapeutic intervention in prostate cancer
CC and other prostate related diseases, in detection, diagnosis, prognosis,
CC prevention, and treatment of prostate cancer (e.g. prostatic
CC intraepithelial neoplasia (PIN), adenocarcinomas, nodular hyperplasia,
CC and large duct carcinomas) and prostate related diseases (e.g. benign
CC prostatic hyperplasia).
XX
SQ Sequence 1603 BP; 531 A; 257 C; 293 G; 522 T; 0 other;

Query Match 27.9%; Score 1456.8; DB 21; Length 1603;
Best Local Similarity 92.8%; Pred. No. 4e-247;

Matches 1590; Conservative 0; Mismatches 2; Indels 122; Gaps 1;			
QY	51	AAGGACACTCTGGCACCAGTTTGGAACTGCGAGTTTAAAGTCTATAAATGAATGAATAA	110
Db	1	AAGGACACTCTGGCACCAGTTTGGAACTGCGAGTTTAAAGTCTATAAATGAATGAATAA	60
QY	111	TGATAGCAAAAGTGGAGGTTTTTAAAGAGCTATTATTAGGTCCTCGGACAGCATCTTTTT	170
Db	61	TGATAGCAAAAGTGGAGGTTTTTAAAGAGCTATTATTAGGTCCTCGGACAGCATCTTTTT	120
QY	171	TCAATTAGGAGCAACCTTTTGGCCTATGCGCTAACCTGTGTGCGAACCTTCCTCTAAT	230
Db	121	TCAATTAGGAGCAACCTTTTGGCCTATGCGCTAACCTGTGTGCGAACCTTCCTCTAAT	180
QY	231	TGGAAATAGTTAAGCAGATTCATAGAGCTCAATGATATAAATTTGCTACTACAGATGCACT	290
Db	181	TGGAAATAGTTAAGCAGATTCATAGAGCTCAATGATATAAATTTGCTACTACAGATGCACT	240
QY	291	GGGACTCAACGTGACCTTTATCAAGTGAGATGGAGTCTTGCCTCTGTCTCAAGGCTGGAGC	350
Db	241	GGGACTCAACGTGACCTTTATCAAGTG	266
QY	351	CCAATGGTGTGATCTTGGCTCACTGCAACCTCCACCTCCAGGTTCAAACTTCTCTCTG	410
Db	267	-----	266
QY	411	CCTCAGCCTCCCAAGTAACCTGGGATTACAGCAGGCTTGGTGCATTTGACACTTCATGATA	470
Db	267	-----AGCAGGCTTGGTGCATTTGACACTTCATGATA	298
QY	471	TCAGCAAAAGTGGAACTAAACACAGCTCCTCGAAGAGGACTATGACATCATCAGTTGGG	530
Db	299	TCATCCAAAGTGGAACTAAACACAGCTCCTCGAAGAGGACTATGACATCATCAGTTGGG	358
QY	531	AGTCTCCAGGACAGCGGACCCCTTTGGAAAGGACTAGAAAGTGTGAATCTATTAGTCT	590
Db	359	AGTCTCCAGGACAGCGGACCCCTTTGGAAAGGACTAGAAAGTGTGAATCTATTAGTCT	418
QY	591	TCGATATCAAAATCTCTGCTCTGTATAAGCATTTTCAATTTACAAGACAGAGCCCTACT	650
Db	419	TCGATATGAAATCTCTGCTCTGTATAAGCATTTTCAATTTACAAGACAGAGCCCTACT	478
QY	651	CCTAGGCGAGCAAAAGTGGCAACAGGCAAGCAGAGGGAAGAGATCATCAGGCAATTC	710
Db	479	CCTAGGCGAGCAAAAGTGGCAACAGGCAAGCAGAGGGAAGAGATCATCAGGCAATTC	538
QY	711	AGAGTGCACTCTCTTTTCATATATTTCTCAATGCCGTATGTTGGTTTATTTGGCCAA	770
Db	539	AGAGTGCACTCTCTTTTCATATATTTCTCAATGCCGTATGTTGGTTTATTTGGCCAA	598
QY	771	GCATAACAATCTGCTCAAGAAAAAATCTGGAGAAACAAAGGTGCTTTGGCAATGT	830
Db	599	GCATAACAATCTGCTCAAGAAAAAATCTGGAGAAACAAAGGTGCTTTGGCAATGT	658
QY	831	TATGTTCTTTTTCAGACAGCCCTGAGATTTCTGAGGGGAATTCACATAAATGGATCAGG	890
Db	659	TATGTTCTTTTTCAGACAGCCCTGAGATTTCTGAGGGGAATTCACATAAATGGATCAGG	718
QY	891	TCATTTCATTTACGTTGTGCAATATGATTTAAAGATACAACTTTGCAGAGAGCATGC	950
Db	719	TCATTTCATTTACGTTGTGCAATATGATTTAAAGATACAACTTTGCAGAGAGCATGC	778
QY	951	TTTCTTAGGCTAGGACGCTGAGGACTTAAGGTAAGGCTTCTCAAGATCAGTTAATC	1010
Db	779	TTTCTTAGGCTAGGACGCTGAGGACTTAAGGTAAGGCTTCTCAAGATCAGTTAATC	838
QY	1011	AAGAAAGTGTCTTTGCAATCTCTGAAATGCCCTTTGTTGCAAAATATTGGTTATATTGATTA	1070
Db	839	AAGAAAGTGTCTTTGCAATCTCTGAAATGCCCTTTGTTGCAAAATATTGGTTATATTGATTA	898
QY	1071	AATTACACTTAAATGGAAACAACTTTTAACCTTACAGATGAACAAACCCACAAAAGCAAAA	1130
Db	899	AATTACACTTAAATGGAAACAACTTTTAACCTTACAGATGAACAAACCCACAAAAGCAAAA	958

QY	1131	AATCAAAAGCCCTACCTATGATTTCAATATTTCTGTGAACCTGGATTAAAGGATTCCTGCG	1190
Db	959	AATCAAAAGCCCTACCTATGATTTCAATATTTCTGTGAACCTGGATTAAAGGATTCCTGCG	1018
QY	1191	TTGCTTTTGGGCATAAATAATGATAATGGAATATTTCCAGGTATTTGTTAAAAATGAGGGCCCA	1250
Db	1019	TTGCTTTTGGGCATAAATAATGATAATGGAATATTTCCAGGTATTTGTTAAAAATGAGGGCCCA	1078
QY	1251	TCTACAAATCTTAGCAATACCTTTGGGATAATTTCAAAATTCAGCTGGACATTTGCTAATTT	1310
Db	1079	TCTACAAATCTTAGCAATACCTTTGGGATAATTTCAAAATTCAGCTGGACATTTGCTAATTT	1138
QY	1311	GTTTTTTATATACATCTTTTGTCTAGAAATTTCAAAATTTAAAGTATGTGAATTTAGTTAAATTA	1370
Db	1139	GTTTTTTATATACATCTTTTGTCTAGAAATTTCAAAATTTAAAGTATGTGAATTTAGTTAAATTA	1198
QY	1371	GCTGTGCTGATCAATTTCAAAACAACTACTTTCTCTAAATTTTAGACTATGAAGGTCATATA	1430
Db	1199	GCTGTGCTGATCAATTTCAAAACAACTACTTTCTCTAAATTTTAGACTATGAAGGTCATATA	1258
QY	1431	TTCAACAAATATATCTACACATACAAATTTATAGATTTGTTTTTCATTTATATATGCTTTCATCT	1490
Db	1259	TTCAACAAATATATCTACACATACAAATTTATAGATTTGTTTTTCATTTATATATGCTTTCATCT	1318
QY	1491	TAACAGAAATCTCTTTTGTGATTGTTTTTAGAAAACTGAGAGTGTTTTAATTCATAATTAATCTT	1550
Db	1319	TAACAGAAATCTCTTTTGTGATTGTTTTTAGAAAACTGAGAGTGTTTTAATTCATAATTAATCTT	1378
QY	1551	GATCAAAAAATTTGGGAACAATCCAGCATTAATTTGATGTGATTTGTTTTTATGTACATA	1610
Db	1379	GATCAAAAAATTTGGGAACAATCCAGCATTAATTTGATGTGATTTGTTTTTATGTACATA	1438
QY	1611	AGGAGTCTTAAGCTTGGTGCCTTGAAGTCTTTTGTACTTAGTCCCATGTTTAAATTAATCT	1670
Db	1439	AGGAGTCTTAAGCTTGGTGCCTTGAAGTCTTTTGTACTTAGTCCCATGTTTAAATTAATCT	1498
QY	1671	ACTTTATATCTAAAGCATTTATGTTTTTCAATTTCAATTTACATGATGCTAATTTATGGCAA	1730
Db	1499	ACTTTATATCTAAAGCATTTATGTTTTTCAATTTCAATTTACATGATGCTAATTTATGGCAA	1558
QY	1731	TTATAACAAATATTAAAGATTTTCAAAATAGATAA	1764
Db	1559	TTATAACAAATATTAAAGATTTTCAAAATAGATAA	1592

RESULT 4

ABV23954/c

ID ABV23954 standard; cDNA; 1804 BP.

AC ABV23954;

XX

DT 16-SEP-2002 (first entry)

XX

XX Human prostate expression marker cDNA 23945.

DE

XX

KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;

XX pharmacogenomic marker; gene; ss.

XX Homo sapiens.

OS

PN WO200160860-A2.

XX

PD 23-AUG-2001.

XX

PF 20-FEB-2001; 2001WO-US05171.

XX

PR 17-FEB-2000; 2000US-183319P.

PR

PR 16-MAR-2000; 2000US-189862P.

PR

PR 25-MAY-2000; 2000US-207454P.

PR

PR 09-JUN-2000; 2000US-211314P.

PR

PR 18-JUL-2000; 2000US-219007P.

PR

PR 13-DEC-2000; 2000US-255281P.

ID ABV24032 standard; cDNA; 1804 BP.
XX AC ABV24032;
XX 16-SEP-2002 (first entry)
XX DE Human prostate expression marker cDNA 24023.
XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX OS Homo sapiens.
XX PN WO200160860-A2.
XX PD 23-AUG-2001.
XX PF 20-FEB-2001; 2001WO-US05171.
XX PR 17-FEB-2000; 2000US-183319P.
XX PR 26-MAR-2000; 2000US-189862P.
XX PR 15-MAY-2000; 2000US-207454P.
XX PR 09-JUN-2000; 2000US-211314P.
XX PR 18-JUL-2000; 2000US-219007P.
XX PR 13-DEC-2000; 2000US-255281P.
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PI Schlegel R, Endege WO, Monahan JE;
XX WPI; 2001-662795/76.
XX PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX Claim 1; Page 4455-4456; 11750pp; English.
XX CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX SQ Sequence 1804 BP; 529 A; 384 C; 333 G; 544 T; 14 other;
Query Match 25.1%; Score 1308; DB 23; Length 1804;
Best Local Similarity 94.7%; Pred No. 6e-221;
Matches 1507; Conservative 0; Mismatches 65; Indels 20; Gaps 14;
QY 175 TTAGGCAGCAACCTTTTGGCCCTATGCCGTACACCTGTGTCTGCAACTTCTCTAAATGGG 234
DB 1801 TTAGGCAGCAACC-ITTTGCCCTATGCCGTACACCTGTGTCTGCAACTTCTCTAA- 1744
QY 235 AAATAGTTAAGCAGATTCATAGAGCTGAATGATATAAATGTACTACGAGATGCATGGGA 294
DB 1743 AAATAGTTAAGCAGATTCATAGAGCTGAATGATATAAATGTACTACGAGATGCATGGGA 1684
QY 295 CTCACGTCACCTTATCAAGTCAGATGGAGTCTGCCCTGTCTCCAAAGCTGAGCCCAA 354
DB 1683 CTCACGTCACCTTATCAAGTCAGATGGAGTCTGCCCTGTCTCCAAAGCTGAGCCCAA 1624
QY 355 TGGTGTGATCTTGGGTCACTGCAACCTCCACCTCCAGGTTCAAAAGCTTCTCCTGCCCTC 414

1623 TGGTGTGATCTTGGGTCACTGCAACCTCCAGCTCCAGGTTCAAAAGCTTCTCCTGCTGG 1564
QY 415 AGCCTCCCAAGTAAGTGGGATTTACAGCAGGCTTGGTGCATTTGACATTCATGATATCAG 474
DB 1563 GACTCAAGCTGACCTTATCAAGTGAGCAGGCTTGGTGCATTTGACATTCATGATATCAG 1504
QY 475 CCAAGTGGAGTAAACACAGCTCTGGAAGAGGACTATGACATCATCATAGGTTGGAGTC 534
DB 1503 CCAAGTGGAGTAAACACAGCTCTGGAAGAGGACTATGACATCATCATAGGTTGGAGTC 1444
QY 535 TCCAGGACAGCGGACCCCTTTGGAAAAGGACTAGAAAAGTGTGAAATCTATTAGTCTTCGA 594
DB 1443 TCCAGGACAGCGGACCCCTTTGGAAAAGGACTAGAAAAGTGTGAAATCTATTAGTCTTCGA 1384
QY 595 TATGAATCTCTGTCTCTGTATAAAGCATTTCTATTTACAGACACAGGCGCTACTCTCTA 654
DB 1383 TATGAATCTCTGTCTCTGTATAAAGCATTTCTATTTACAGACACAGGCGCTACTCTCTA 1324
QY 655 GGGCAGCAAAAAGTGGCAACAGCAGGAGGAAAAGAGATCATGAGGCATTTTCAGAG 714
DB 1323 GGGCAGCAAAAAGTGGCAACAGCAGGAGGAAAAGAGATCATGAGGCATTTTCAGAG 1264
QY 715 TGCACCTGTCTTTTCATATATTTCTCAATGCCGTATGTTTGGTTTATTTTGGCCCAAGCAT 774
DB 1263 TGCACCTGTCTTTTCATATATTTCTCAATGCCGTATGTTTGGTTTATTTTGGCCCAAGCAT 1204
QY 775 AACATCTGCTCAAGAAAATAAATCTGGAGAAAACAAGGTGCCCTTGGCCAAATGTTATG 834
DB 1203 AACATCTGCTCA--AAAAAAAATCTGGAGAAAACAAGGTGCCCTTGGCCAAATGTTATG 1146
QY 835 TTTCTTTTGGACAAGCCCTGAGATTTCTGAGGGGAATTCACATAAATGGGATCAGTCAAT 894
DB 1145 TTTCTTTTGGACAAGCCCTGAGATTTCTGAGGGGAATTCACATAAATGGGATCAGTCAAT 1086
QY 895 TCATTATCGTGTGTGCAAAATATGATTTAAAGATACAACTTTGCGAGAGAGCATGCTTTC 954
DB 1085 TCATTATCGTGTGTGCAAAATATGATTTAAAGATACAACTTTGCGAGAGAGCATGCTTTC 1026
QY 955 CTAAGGTAGGACGCTGGAGGACTAAGGTAAAGCATTTCTCAAGATCAGTTAAACAAGA 1014
DB 1025 CTAAGGTAGGACGCTGGAGGACTAAGGTAAAGCATTTCTCAAGATCAGTTAAACAAGA 966
QY 1015 AAGGTGCTCTTTGCAATTTCTGAAATGCCCTTGTGCAAAATATGTTTATATGTTAAAT 1074
DB 965 AAGGTGCTCTTTGCAATTTCTGAAATGCCCTTGTGCAAAATATGTTTATATGTTAAAT 906
QY 1075 TACACTTAATGGAACAAACCTTTAACTTTACAGATGAACAAA-CCACAAAAGCAAAAAT 1133
DB 905 TACACTTAATGGAACAAACCTTTAACTTTACAGATGAACAAAAGCAAAAAGCAAAAAT 846
QY 1134 CAAAAGCCCTACCTATGATTTTCATATTTCTGTAACCTGGAATTAAGGATTCCTGCTTG 1193
DB 845 CAAAAGCCCTACCTATGATTTTCATATTTCTGTAACCTGGAATTAAGGATTCCTGCTTG 786
QY 1194 CTTTGGGCAATAATGGAATTAATTCAGGATTTTCCAGGTATTTAAATGAGGCCCATCT 1253
DB 785 CTTTGGGCAATAATGGAATTAATTCAGGATTTTCCAGGTATTTAAATGAGGCCCATCT 726
QY 1254 ACAATCTTACCAATACCTTTGGATAAATTTCTAAATTCAGCTGGACATTTCTTAATGTT 1313
DB 725 ACAATCTTACCAATACCTTTGGATAAATTTCTAAATTCAGCTGGACATTTCTTAATGTT 666
QY 1314 TTTTATATACATCTTGTGTAGAAATTTCAATTTTAAGTATGTAATTTAGTTAAATTTAGC 1372
DB 665 TTTTATATACATCTTGTGTAGAAATTTCAATTTTAAGTATGTAATTTAGTTAAATTTAGC 606
QY 1373 TGTGCT-GATCAATTCAAAACATTTACTTTCTTAAATTTTACATATGAAGGTCAATAAT 1431
DB 605 TGTGTTGATCAATTCAAAACATTTACTTTCTTAAATTTTACATATGAAGGTCAATAAT 546
QY 1432 TCAACAAATATATCTACACATCAATATATAGAT-TGTTTTTTCATTTAATATGTTCTCATCT 1490

Db 545 TCAACAAATATATCTACACATACAAATATAGATGTGTTTTTCATTTATATATATCTTCATCT 486
Qy 1491 TAACAGAAATGCTCTTCTGATGTTTGTAGAAACTGAGAGTTTATATTCATAATTAATCT 1550
Db 485 TAACAGAAATGCTCTTCTGATGTTTGTAGAAACTGAGAGTTTATATTCATAATTAATCT 427
Qy 1551 GATCAAAAAA--TTGTGGGAACATCCAGCATTAAT--TGTATGTGATGTTTTTATGTT 1605
Db 426 GATCAAAAAAATGTTGGGAACAAATCCAGCATTAATTTGTATGTGATGTTTATGTT 367
Qy 1606 ACATAAGAGTCTTAAGC--TTGTGTCCTTGAAGTCTTTTGTACTTATGTCCTCCAGTTAA 1663
Db 366 ACATAAGAGTCTTAAAGCTGGGTGCTTGAAGTCTTTTGTACTTATGTCCTCCAGTTAA 307
Qy 1664 AATTACTACTTATATCTAAGCATTTATGTTTTCATTTCAATTTACATGACTTAAT 1723
Db 306 AATTACTA--TTTATATCTAAGCATTTATGTTTTCATTTCAATTTACATGACTTAAT 250
Qy 1724 ATGCAATTTAACAATAATTAAGATTTTGA 1755
Db 249 AT-GCCATTATAACAATAATTAAGGATCTCAA 219
RESULT 6
ABV29837/c
ID ABV29837 standard; cdna; 1804 BP.
XX
AC ABV29837;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cdna 29828.
XX
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN W0200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 6426; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;

CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ Sequence 1804 BP; 529 A; 384 C; 333 G; 544 T; 14 other;
Query Match 25.18; Score 1308; DB 23; Length 1804;
Best Local Similarity 94.7%; Pred. No. 5e-221;
Matches 1507; Conservative 0; Mismatches 65; Indels 20; Gaps 14;
Qy 175 TTAGGCAGCAACCTTTTGGCCCTATGCCGTAACCTGTCTGCAACTTCTCTCTAATGGG 234
Db 1801 TTAGGCAGCAACCTTTTGGCCCTATGCCGTAACCTGTCTGCAACTTCTCTCTAATGGG 1744
Qy 235 AATATAGTTAAGCAGATTCATAGAGCTGAATGATGATAAATTTGTACTACGAGATGCTACTGGGA 294
Db 1743 AATATAGTTAAGCAGATTCATAGAGCTGAATGATGATAAATTTGTACTACGAGATGCTACTGGGA 1684
Qy 295 CTCAACGTGACCTTATCAAGTGAGATGGAGTCTTGGCCCTGTCTCCAGGCTGGAGCCCAA 354
Db 1683 CTCAACGTGACCTTATCAAGTGAGATGGAGTCTTGGCCCTGTCTCCAGGCTGGAGCCCAA 1624
Qy 355 TGGTGTGATCTTGGCTCACCTGCAACCTCCACCTCCAGGTTCAAAACGTTTCTCTGCTC 414
Db 1623 TGGTGTGATCTTGGCTCACCTGCAACCTCCACCTCCAGGTTCAAAACGTTTCTCTGCTC 1564
Qy 415 AGCCTCCCAAGTAACTGGGATTACAGCAGGCTTGGTGCATTTGACACTTTCATGATATCAG 474
Db 1563 GACTCAACGTGACCTTATCAAGTGAGCAGGCTTGGTGCATTTGACACTTTCATGATATCAG 1504
Qy 475 CCAAGTGGAACTTAAACACAGCTCTGGAAGAGACTATGACATCATCAGTGGGAGTC 534
Db 1503 CCAAGTGGAACTTAAACACAGCTCTGGAAGAGACTATGACATCATCAGTGGGAGTC 1444
Qy 535 TCCAGGACAGCGGACCCCTTTGGAAAAGGACTAGAAAGTGTGAAATCTATTAGTCTTCGA 594
Db 1443 TCCAGGACAGCGGACCCCTTTGGAAAAGGACTAGAAAGTGTGAAATCTATTAGTCTTCGA 1384
Qy 595 TATGAAATCTCTGTCTCTGTAAAGCATTTTCATATTTTACAAGACACAGGCTTCTCTCTA 654
Db 1383 TATGAAATCTCTGTCTCTGTAAAGCATTTTCATATTTTACAAGACACAGGCTTCTCTCTA 1324
Qy 655 GGCAGCAAAAAGCTGCAACAGCAAGCAGGAGGAAAGAGATCATGAGGCAATTCAGAG 714
Db 1323 GGCAGCAAAAAGCTGCAACAGCAAGCAGGAGGAAAGAGATCATGAGGCAATTCAGAG 1264
Qy 715 TGCACCTCTCTTTTCATATATTTCTCAATGCCGTATGTTGGTGTGTTTATTTTGGCCAAAGCAT 774
Db 1263 TGCACCTCTCTTTTCATATATTTCTCAATGCCGTATGTTGGTGTGTTTATTTTGGCCAAAGCAT 1204
Qy 775 AACAACTCTCTCAAGAAAAAATACTGGAGAAACAAAGGTGCTTTTGGCAATGTTATG 834
Db 1203 AACAACTCTCTCA--AAAAAATACTGGAGAAACAAAGGTGCTTTTGGCAATGTTATG 1146
Qy 835 TTTCTTTTTCACAGCCCTGAGATTTCTGAGGGAATTCACATAAATGGATCAGGTCAT 894
Db 1145 TTTCTTTTTCACAGCCCTGAGATTTCTGAGGGAATTCACATAAATGGATCAGGTCAT 1086
Qy 895 TCATTTACGTTGTGTGCAAAATATGATTTAAAGATFACAAACCTTTGAGAGAGCATGCTTTC 954
Db 1085 TCATTTACGTTGTGTGCAAAATATGATTTAAAGATFACAAACCTTTGAGAGAGCATGCTTTC 1026
Qy 955 CTAAAGGTAGGAGCTGGAGACTAAGGTTAAAGCATTTCTCAAGATCAGTTAATCAAGA 1014
Db 1025 CTAAAGGTAGGAGCTGGAGACTAAGGTTAAAGCATTTCTCAAGATCAGTTAATCAAGA 966
Qy 1015 AAGTGTCTTTTGCATTTCTGAAATGCCCTTGTGTCAAATATGTTTATTTATGATTAAT 1074
Db 965 AAGTGTCTTTTGCATTTCTGAAATGCCCTTGTGTCAAATATGTTTATTTATGATTAAT 906
Qy 1075 TACACTTAAATGGAACAAACCTTTTAACTTACAGATGAACAAA-CCACAAAAAGCAAAAAT 1133

Db 905 TACACTTAATGGAACAACCTTTAACTTACAGATGAACAACCCACAAAAGCAAAAAAT 846
QY 1134 CAAAAGCCCTACCTATGATTTTCATATTTCTGTGTAAGTGAATTAAGGATTCCTGCTTG 1193
Db 845 CAAAAGCCCTACCTATGATTTTCATATTTCTGTGTAAGTGAATTAAGGATTCCTGCTTG 786
QY 1194 CTTTGGGCATAAATGATAATGGAATATTTCCAGGTATTTTAAATGAGGCGCCATCT 1253
Db 785 CTTTGGGCATAAATGATAATGGAATATTTCCAGGTATTTTAAATGAGGCGCCATCT 726
QY 1254 ACAATTTCTAGCAATCTTGGGATAATTCATAATTCAGTGGACATTTGCTTAATGTT 1313
Db 725 ACAATTTCTAGCAATCTTGGGATAATTCATAATTCAGTGGACATTTGCTTAATGTT 666
QY 1314 TTTTATATACATCTTGTGTAAGTGAATTTCAAAATTTAAGTATGTAATTTAGTTAA--TTAGC 1372
Db 665 TTTTATATACATCTTGTGTAAGTGAATTTCAAAATTTAAGTATGTAATTTAGTTAA--TTAGC 606
QY 1373 TGTGCT-GATCAATTCAAAACATTTACTTTCTCTAAATTTTAACTATGAAGTGCATAAAT 1431
Db 605 TGTGTTGGATCAATTCAAAACATTTACTTTCTCTAAATTTTAACTATGAAGTGCATAAAT 546
QY 1432 TCACAAATATATCTACACATACAAATATAGAT-TGTTTTTCATTAATGCTTTTCATCT 1490
Db 545 TCACAAATATATCTACACATACAAATATAGATGTTTTCATTAATGCTTTTCATCT 486
QY 1491 TAACAGAATTCGTTTGTGATTTTGTAGAAACTGAGAGTTTTAAATTCATAATTAATTT 1550
Db 485 TAACAGAATTCGTTTGTGATTTTGTAGAAACTGAGAGTTTTAAATTCATAATTAATTT 427
QY 1551 GATCAAAAA--TTGTGGGAACAATCCAGCATTAAT---TGTATGTGATTTGTTTATGT 1605
Db 426 GATCAAAAAATGTTGGGAACAATCCAGCATTAATTTGTATGTTGTTTATGTT 367
QY 1606 ACATAAGAGTCTTTAAGC--TTGGTGCCTTGAAGTCTTTTGTACTTTAGTCCCATGTTTAA 1663
Db 366 ACATAAGAGTCTTTAAGTGTGGTGCCTTGAAGTCTTTTGTACTTTAGTCCCATGTTTAA 307
QY 1664 AATTACTACTTTATATCTAAGCATTTTATGTTTTCATTTTCAATTTACATGATGCTAAT 1723
Db 306 AATTACTA-TTTATATCTAAGCATTTTATGTTT--CAATTCAATTACATGATGTTAAT 250
QY 1724 ATGGCAATTAATAACAATATTAAGATTTTCA 1755
Db 249 AT-GCCATTATAACAATATTAAGGATCTCAA 219

RESULT 7

ABV29916/c
ID ABV29916 standard; cDNA; 1804 BP.
XX AC ABV29916;
XX DT 16-SEP-2002 (first entry)
XX DE Human prostate expression marker cDNA 29907.
XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX KW pharmacogenomic marker; gene; ss.
XX OS Homo sapiens.
XX PN WO200160860-A2.
XX XX 23-AUG-2001.
XX PD 20-FEB-2001; 2001WO-US05171.
XX PF 17-FEB-2000; 2000US-183319P.
XX PR 16-MAR-2000; 2000US-189862P.
XX PR 25-MAY-2000; 2000US-207454P.
XX PR 09-JUN-2000; 2000US-211314P.

18-JUL-2000; 2000US-219007P.
13-DEC-2000; 2000US-255281P.
(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
Schlegel R, Endege WO, Monahan JE;
WPI; 2001-662795/76.
Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer -
Claim 1; Page 6454; 11750pp; English.
The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for:
(a) assessing whether a patient is afflicted with prostate cancer;
(b) monitoring the progression of prostate cancer in a patient;
(c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient;
(d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient;
(e) selecting a composition for inhibiting prostate cancer in a patient;
(f) assessing the prostate cell carcinogenic potential of a compound;
(g) determining whether prostate cancer has metastasized in a patient;
(h) assessing the aggressiveness or indolence of prostate cancer in a patient;
(I) is also useful as a pharmacodynamic or pharmacogenomic marker.
Sequence 1804 BP; 529 A; 384 C; 333 G; 544 T; 14 other;
Query Match 25.1%; Score 1308; DB 23; Length 1804;
Best Local Similarity 94.7%; Pred. No. 6e-221;
Matches 1507; Conservative 0; Mismatches 65; Indels 20; Gaps 14;
QY 175 TTAGCAGCAACCTTTTGGCCCTATGCGGTAACTGTCTGCAACTTCTCTTAATTTGGG 234
Db 1801 TTAGCAGCAAC--TTTGGCCCTATGCGGTAACTGTCTGCAACTTCTCTTAAT-TGGG 1744
QY 235 AAATAGTTAAGCAGATTCATAGAGTGAATGATAAAATTTGACTACGAGATGCTAGGGA 294
Db 1743 AAATAGTTAAGCAGATTCATAGAGTGAATGATAAAATTTGACTACGAGATGCTAGGGA 1684
QY 295 CTCACGTGACCTTTATCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 354
Db 1683 CTCACGTGACCTTTATCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1624
QY 355 TGGTGTGATCTTGGCTCACTCAACCTCCACCTCCAGGTTTCAACGTTTCTCTGCTC 414
Db 1623 TGGTGTGATCTTGGCTCACTCAACCTCCACCTCCAGGTTTCAACGTTTCTCTGCTC 1564
QY 415 AGCCTCCCAAGTAACTGGGATTTACAGAGGCTTGGTGCATTTGACATTCATGATATCAG 474
Db 1563 GACTCAACGTGACCTTTATCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1504
QY 475 CCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 534
Db 1503 CCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1444
QY 535 TCCAGGACAGGACCTTTTGGAAAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 594
Db 1443 TCCAGGACAGGACCTTTTGGAAAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1384
QY 595 TATGAATTCCTGCTCTGTCTGTAAGAGCATTTTCAATTTTACAAGACAGGCTACTCCTA 654
Db 1383 TATGAATTCCTGCTCTGTCTGTAAGAGCATTTTCAATTTTACAAGACAGGCTACTCCTA 1324
QY 655 GGGCAGCAAAAAGTGGCAACAGGCAAGCAGAGGGAAGAGATCATGAGGCAATTTTCAGAG 714
Db 1323 GGGCAGCAAAAAGTGGCAACAGGCAAGCAGAGGGAAGAGATCATGAGGCAATTTTCAGAG 1264

Db 152 ----- 151
QY 406 TCCTGCCTCAGCCTCCCAAGTAACCTGGGATACACAGCCTTGGTGCAATTCACACTTCA 465
Db 152 -----ACAGGCTTGGTGCAATTCACACTTCA 178
QY 466 TGATATCAGCCAAAGTGAACCTAAACAGCTCCTCGAAGAGGACTATGACATCATCAGG 525
Db 179 TGATATCATCCAAAGTGAACCTAAACAGCTCCTCGAAGAGGACTATGACATCATCAGG 238
QY 526 TTGGAGTCTCCAGGACAGCGGCCCTTTGGAAAAGGACTAGAAAAGTGTGAATCTATT 585
Db 239 TTGGAGTCTCCAGGACAGCGGCCCTTTGGAAAAGGACTAGAAAAGTGTGAATCTATT 298
QY 586 AGCTTTCGATATGAATCTCTGCTCTGTAAGCAATTCATATTTACAAGACACAGC 645
Db 299 AGCTTTCGATATGAATCTCTGCTCTGTAAGCAATTCATATTTACAAGACACAGC 358
QY 646 CTACTCTTAGGSCAGCAAAAAGTGGCAACAGCAGAGGAGGAAAAGAGATCATGAGGC 705
Db 359 CTACTCTTAGGSCAGCAAAAAGTGGCAACAGCAGAGGAGGAAAAGAGATCATGAGGC 418
QY 706 ATTTCCAGAGTGCACGTCTTTTCATATATTTCTCAATGCCGTATGTTGGTTTATTTTG 765
Db 419 ATTTCCAGAGTGCACGTCTTTTCATATATTTCTCAATGCCGTATGTTGGTTTATTTTG 478
QY 766 GCCAAGCATAACAATCTGCTCAAGAAAAAATCTGAGAAAAAAGGTCCTTGC 825
Db 479 GCCAAGCATAACAATCTGCTCAAGAAAAAATCTGAGAAAAAAGGTCCTTGC 538
QY 826 AATGTTATGTTTCTTTTGACAAGCCCTGAGATTTCTGAGGGAATTCACATAAATGGGA 885
Db 539 AATGTTATGTTTCTTTTGACAAGCCCTGAGATTTCTGAGGGAATTCACATAAATGGGA 598
QY 886 TCAGTCAATCAATTTACGTGTGTGCAAAATGATTTAAAGATACAACTTTGCAGAGAG 945
Db 599 TCAGTCAATCAATTTACGTGTGTGCAAAATGATTTAAAGATACAACTTTGCAGAGAG 658
QY 946 CATGCTTTCCTAAGGTTAGGACGTGGAGGACTAAGGTTAAGCATTTCTCAAGATCAGT 1005
Db 659 CATGCTTTCCTAAGGTTAGGACGTGGAGGACTAAGGTTAAGCATTTCTCAAGATCAGT 718
QY 1006 TAATCAAGAAAGTGCTCTTTGCAATTCGAAATGCCCTTTGTCAAATATTTGTTATATT 1065
Db 719 TAATCAAGAAAGTGCTCTTTGCAATTCGAAATGCCCTTTGTCAAATATTTGTTATATT 778
QY 1066 GATTAATTTACACTTAATGGAAACAACCTTTAACTTACAGATGAACAAACCCACAAG 1125
Db 779 GATTAATTTACACTTAATGGAAACAACCTTTAACTTACAGATGAACAAACCCACAAG 838
QY 1126 CAAAATAACAAAGCCCTACCTATGATTTTCATATTTCTGTGTAAGTAAAGGATT 1185
Db 839 CAAAATAACAAAGCCCTACCTATGATTTTCATATTTCTGTGTAAGTAAAGGATT 898
QY 1186 CCGTCTGCTTTTGGGCATAAATGATAATGGAATATTTCCAGTATTTGTTTAAATGAGG 1245
Db 899 CCGTCTGCTTTTGGGCATAAATGATAATGGAATATTTCCAGTATTTGTTTAAATGAGG 958
QY 1246 GCCCATCTACAAATTTCTAGCAATACTTTGGATAATTTCTAAATTCAGCTGGACATGTGC 1305
Db 959 GCCCATCTACAAATTTCTAGCAATACTTTGGATAATTTCTAAATTCAGCTGGACATGTGC 1018
QY 1306 TAAATGTT 1312
Db 1019 TAAATGTT 1025

RESULT 9
ABZ80229/c
ID ABZ80229 standard; DNA; 249999 BP.
XX
AC ABZ80229;
XX

DT 02-JUN-2003 (first entry)
XX Human tramdorin gene region genomic DNA SEQ ID NO:26.
XX
KW Neuroprotective; nootropic; cerebroprotective; analgesic; gene therapy;
central nervous system disorder; CNS disorder; multiple sclerosis;
nerve injury; neuropathic pain; stroke; trauma; non-CNS disorder; tram;
tramdorin; human; chromosome 5; gene; ds.
XX
OS Homo sapiens.
XX WO2003016502-A2.
XX 27-FEB-2003.
XX 21-AUG-2002; 2002WO-US26637.
XX 21-AUG-2001; 2001US-313907P.
XX 21-AUG-2002; 2002US-0225810.
XX (MCLA-) MCLAUGHLIN RES INST.
XX Birmingham JR;
XX WPI; 2003-278567/27.
XX New nucleic acid sequence encoding tramdorins, e.g. mouse tramd 1,
mouse tramd 2, mouse tramd 3, human tramd 1, human tramd 2, human tramd
3 or rat tramd 1, useful for treating CNS, e.g. stroke, multiple
sclerosis, trauma, neuropathic pain -
XX Example 6; Fig 9; 177pp; English.
XX The present invention describes an isolated nucleic acid sequence
comprising a cDNA sequence encoding mouse tramdorin (tramd) 2, mouse
tramd 3, human tramd 1, human tramd 2, human tramd 3 or rat tramd 1, or
the genomic sequence of mouse tramd 1 or mouse tramd 3. Mouse tramd 1 is
located to chromosome 11, whereas human tramd 1 is located to chromosome
5q31-33. The tramd sequences have neuroprotective, nootropic, analgesic
and cerebroprotective activities, and can be used in gene therapy. The
nucleic acid sequences are useful for diagnosing and treating central
nervous system (CNS) disorders such as multiple sclerosis, nerve injury,
neuropathic pain, stroke or trauma, and non-CNS disorders. The present
sequence represents the genomic sequence of the human tramdorin gene
region, which is given in the exemplification of the present invention.
XX
SQ Sequence 249999 BP; 75050 A; 54012 C; 51931 G; 67805 T; 1201 other;
Query Match 8.9%; Score 462; DB 25; Length 249999;
Best Local Similarity 84.7%; Pred. No. 5,7e-72;
Matches 574; Conservative 0; Mismatches 80; Indels 24; Gaps 4;
QY 3363 AGTGTGCCCGGAATTTATTTCTTCTGGTGGGTTCTTGGTCTCGCTGACTTCAAGAATGAA 109919
Db 109978 AGTGTGTCCGGAATTTGGTTCTTCTGGTGGGTTCTTGGTCTCGCTGACTTCAAGAATGAA 3422
QY 3423 ACCGACGACCCCTTGAGTGCAGTGTACAGTCTTAAAGATGCTGTCTCAGAGTTTGTTC 3482
Db 109918 GCCAGGCGCCCTCGCGGTGAGTGTACAGTCTTAAAGATGCTGTCTCGCGAGTTGTTC 109859
QY 3483 CTTTCAGATGTTTCAGATGTCTCGGAGTGTTCCTCCCTATGTTGGTGTCTGCTCGCTGA 3542
Db 109858 CTTTCAGATGTTTCAGATGTCTCGGAGTGTTCCTCCCTCTGGTGGGTTCTGCTGCTGA 109799
QY 3543 CTTTCAGATGTTTCAGATGTCTCGGAGTGTTCCTCCCTATGTTGGTGTCTGCTCGCTGA 3602
Db 109798 CTTTCAGAGTGAAGCTGCAGACCTTTGCAGTGA--GTGTTACAGCTCTTAAAGTGGGCGC 109741
QY 3603 GTCCAGAGTGTCTTCTTCTCCCGGTAGTGTCTCGCTGATGTCAGCAATGAAGC 3662
Db 109740 GTCCAGAGTGTCTTCTTCTCCCGGTAGTGTCTCGCTGATGTCAGCAATGAAGC 109681
QY 3663 TGCAGACCCCTCGCGGTAAGTGTTCAGAGCTCATAAAGGTAGTGCAAAACCCACAGTGAGC 3722


```
RESULT 11
AB089916
ID AB089916 standard; CDNA; 843 BP.
XX
AC AB089916;
XX
DT 27-SEP-2002 (first entry)
XX
DE Human prostate expressed polynucleotide SEQ ID NO 1172.
XX
KW Human; prostate; cytostatic; tumour; cancer; vaccine; gene therapy;
KW gene; ss.
XX
OS Homo sapiens.
XX
PN WO200255700-A2.
XX
PD 18-JUL-2002.
XX
PF 07-DEC-2001; 2001WO-US47349.
XX
PR 07-DEC-2000; 2000US-254648P.
XX
PR 13-MAR-2001; 2001US-275688P.
XX
PA (CHIR ) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Escobedo J, Garcia PD, Kassam A, Lamson G, Drmanac R;
PI Krkenjakov R, Dickinson M, Drmanac S, Labat I, Leshkowitz D;
PI Kita D, Garcia V, Jones WL, Stache-Crain B, Scott EM;
XX
WPI; 2002-557824/59.
XX
New genes and gene products isolated from human prostate, useful for
treating or diagnosing tumor or cancer (e.g. prostate cancer or breast
cancer), or as vaccines for treating or preventing these diseases -
XX
Claim 1; SEQ ID NO 1172; 186pp + Sequence Listing; English.
XX
The invention relates to an isolated polynucleotide comprising any of
1477 sequences or its fragment, degenerate variant, antisense or
complement. The polynucleotides and gene products are useful for treating
or diagnosing tumor or cancer (e.g. prostate cancer, breast cancer,
lung cancer or medullary carcinoma) in a subject (e.g. cattle, dogs,
cats, rabbits, horse or human). The polynucleotides and polypeptides are
also useful as vaccines for treating or preventing these diseases. The
polynucleotides are useful for gene therapy. The present sequence is that
of one of a group of polynucleotides (AB088745-AB090015) disclosed
electronically as sequences of the invention. However only 1271
polynucleotide sequences are given, whereas 1477 polynucleotides and 91
proteins are claimed.
XX
Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequence.
XX
Sequence 843 BP; 214 A; 190 C; 171 G; 251 T; 17 other;
XX
Query Match 8.4%; Score 438.4; DB 24; Length 843;
Best Local Similarity 99.8%; Pred. No. 4.6e-68;
Matches 439; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GAAACTTTAAATATCCCTCAGTGTCTCTGTTAATTCATGTGTAGTGCCTCCAGGCACCTCT 60
Db |||||||
QY 61 GGCACCCAGTTTGGAACTGCAGTTTTTAAAGTCATAAATGAATGAATAATGATACAAA 120
Db |||||||
QY 183 GGCACCCAGTTTGGAACTGCAGTTTTTAAAGTCATAAATGAATGAATAATGATACAAA 242
QY 121 GTTGAGAGCTTTTAAAGAGCTATTATAGTCCCTGGACAGCATCTTTTCAATTAGGC 180
Db |||||||
QY 243 GGTGAGGTTTTTAAAGAGCTATTATAGTCCCTGGACAGCATCTTTTCAATTAGGC 302
```

```
QY 181 AGCAACCTTTTGGCCTATGCCGTAACTGTGTCTGCAACTTCTCTAATTGGGAAATAG 240
Db |||||||
QY 303 AGCAACCTTTTGGCCTATGTCGTAACCTGTCTGCAACTTCTCTAATTGGGAAATAG 362
Db |||||||
QY 241 TTAAGCAGATTTCATAGAGCTGAATGATATAAATTGTACTACGAGATGCCTGGGACTCAAC 300
Db |||||||
QY 363 TTAAGCAGATTTCATAGAGCTGAATGATATAAATTGTACTACGAGATGCCTGGGACTCAAC 422
QY 301 GTGACCTTTATCAAGTGAGATGGAGTCTTGGCCCTGTCTCAAGGCTGGAGCCCAATGGTGT 360
Db |||||||
QY 423 GTGACCTTTATCAAGTGAGATGGAGTCTTGGCCCTGTCTCAAGGCTGGAGCCCAATGGTGT 482
QY 361 GATCTTGGCTCACTGCAACCTCCACCTCCACAGTTTCAACAGTTTCTCTGCTCAGCCTC 420
Db |||||||
QY 483 GATCTTGGCTCACTGCAACCTCCACCTCCACAGTTTCAACAGTTTCTCTGCTCAGCCTC 542
QY 421 CCAAGTAAGTGGGATTACAG 440
Db |||||||
QY 543 CCAAGTAAGTGGGATTACAG 562
Db |||||||
RESULT 12
ABN95882/c
ID ABN95882 standard; DNA; 33206 BP.
XX
AC ABN95882;
XX
DT 13-AUG-2002 (first entry)
XX
DE Gene #2380 used to diagnose liver cancer.
XX
KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW metastatic liver tumour; cytostatic; expression profile; disease state;
KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS Homo sapiens.
XX
PN WO200229103-A2.
XX
PD 11-APR-2002.
XX
PF 02-OCT-2001; 2001WO-US30589.
XX
PR 02-OCT-2000; 2000US-237054P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
WPI; 2002-426119/45.
XX
Diagnosing and detecting the progression of liver cancer,
hepatocellular carcinoma or metastatic liver tumor in a patient,
involves detecting the level of expression of two or more genes in a
liver tissue sample -
XX
Claim 1; SEQ ID NO 2380; 298pp; English.
XX
The invention relates to a novel method for diagnosing and detecting the
progression of liver cancer, hepatocellular carcinoma or metastatic liver
tumour in a patient, and differentiating metastatic liver cancer from
hepatocellular carcinoma in a patient, involving detecting the level of
expression of two or more genes represented in ABN93503-ABN97455 in a
tissue sample. The method of the invention has hepatotropic, and
cytostatic activity. The method is useful for diagnosing and detecting
the progression of liver cancer, hepatocellular carcinoma and metastatic
liver carcinoma in a patient. The method is useful for identifying
expression profiles which serve as useful diagnostic markers as well as
markers that can be used to monitor disease states, disease progression,
drug toxicity, drug efficacy and drug metabolism.
XX
Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.
CC
```

[illegible]

RESULT 13

AAC74309
ID AAC74309 standard; cDNA; 1761 BP.

XX
AC AAC74309.

DT 02-FEB-2001 (first entry)

XX DE Human secreted protein gene 30 SEQ ID NO:40.

XX Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
 KW antirheumatic; antiproliferative; cytostatic; cardiac; vasotropic;
 KW cerebroprotective; neotrophic; neuroprotective; antibacterial; virucide;
 KW fungicide; ophthalmological; veterinary; gene therapy; angiogenesis;
 KW autoimmune disease; hyperpliferative disorder; infection; skin aging;
 KW wound healing; cardiovascular disorder; cerebrovascular disorder;
 KW nervous system disorder; food additive; preservative; ss;
 KW

XX	Homo sapiens.
OS	
XX	WO2000057903-A2.
PN	
XX	
PD	05-OCT-2000.
XX	
XX	22-MAR-2000; 2000WO-US07525.
PF	
XX	26-MAR-1999; 99US-0126595.
PR	
XX	22-DEC-1999; 99US-0171549.
PR	
XX	(HUMA-) HUMAN GENOME SCI INC.
PA	
XX	
PI	Rosen CA, Ruben SM, Komatsoulis G;
XX	
DR	WPI: 2000-594630/56.
DR	P-PSDB; AAB39339.
XX	
PT	New nucleic acid molecules encoding 48 human secreted proteins for
PT	diagnosing, preventing, treating or ameliorating medical conditions and
PT	used as food additives or preservatives -
XX	
XX	Claim 1; Page 338; 395pp; English.
XX	
CC	The polynucleotide sequences given in AAC74280 to AAC74327 encode the
CC	human secreted proteins given in AAB39310 to AAB39357. AAB39358 to
CC	AAB39400 represent human secreted polypeptide sequences and proteins
CC	homologous to them, which are given in the exemplification of the present
CC	invention. Human secreted proteins have activities based on the tissues
CC	and cells the genes are expressed in. Examples of activities include:
CC	antiarthritic; immunosuppressive; antiflammatic; antiproliferative;
CC	cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
CC	neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
CC	and vulnerary. The polynucleotides and polypeptides can be used to
CC	prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC	rabbits, goats, horses, cats, dogs, chickens or sheep. They can also used
CC	in diagnosing a pathological condition or susceptibility to a
CC	pathological condition. Disorders which are diagnosed or treated include
CC	autoimmune diseases, hyperproliferative disorders, cardiovascular
CC	disorders, cerebrovascular disorders, angiogenesis, nervous system
CC	disorders, infections caused by bacteria, viruses and fungi and ocular
CC	disorders. The polypeptides can also be used to aid wound healing and
CC	epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC	maintain organs before transplantation, for supporting cell culture of
CC	primary tissues, to regenerate tissues and in chemotaxis. The
CC	polypeptides can also be used as a food additive or preservative to
CC	increase or decrease storage capabilities, fat content, lipid, protein,
CC	carbohydrate, vitamins, minerals, cofactors and other nutritional
CC	components. AAC74271 to AAC74279 and AAB39309 represent sequences used in
CC	the exemplification of the present invention.
XX	
SQ	Sequence 1761 BP: 409 A; 406 C; 439 G; 507 T; 0 other;

Query Match 8.0%; Score 417.6; DB 21; Length 1761;
Best Local Similarity 82.7%; Pred. No. 2.3e-64;
Matches 516; Conservative 0; Mismatches 99; Indels 9; Gaps 3;

3363	AGTGTCCCGGAATTTATTTCTCTCGTGGTTCCTGGTCTCGCTGACCTCAAGAATGAA	3422
Qy		
702	ATTATCTGGAATTGGTCTCTCCGAGGGTCTTTGGTCTCGCTGACITCAAGAATGAA	761
Db		
3423	ACCGCAGACCCTTGAGGTCAGTCTCAAGTCTTTAAAGATGGTGTTCAGAGTTTGTTC	3482
Qy		
762	GCCATGGACCCCTCGCAGTGAAGTGTTCCTTAAAGATGGTGTTCGCGGAGTTTGTGT	821
Db		
3483	CTTTCAGATGTCAGATGTGTC - - - CGAGGTTTCTCCCTTATGGTGAAGTTCGTGGTCTCGC	3539
Qy		
822	TTTCAGATGTTTCAGATGTGTCTGGAGAGGTTTCTCCTTCTGGTGGGCTCGTGGTCTCGC	881
Db		
3540	TGACTTCAACAATGAAGCCGCAGACCTTTTGCAGTGAAGTGTGTGACAGTTCTTTAAAGCCAG	3599
Qy		
882	TGACTTCAGGAGTGAAGCCACAGACCTTCACAGTGA - - GCATTACAGCTTTAAAGGTGG	939
Db		

AAV83939/c
ID AAV83939 standard; DNA: 80595 BP.
XX AC AAV83939;
XX
XX 03-MAR-1999 (first entry)
XX
XX HC-contig derived from normal human chromosome 10q25.2 region.
XX
XX Yeast artificial chromosome; YAC; probe: eukaryotic chromosome:
KW neocentromere; replication; extra-chromosomal element; segregation;
KW cell division; artificial chromosome; gene therapy; mardel(10);
KW human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
XX
XX Homo sapiens.
XX
XX W09851790-A1.
XX
XX 19-NOV-1998.
XX
XX 13-MAY-1998; 98WO-AU00352.
XX
XX 26-AUG-1997; 97AU-0008791.
XX
XX 13-MAY-1997; 97AU-0006784.
XX
XX (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
XX Cancilla MR, Choo K, Du Sart D;
XX WPI; 1999-009773/01.
XX
XX New isolated nucleic acid comprising neocentromere sequences from
PT eukaryotic chromosome - used to produce replicable, segregating
PT artificial chromosomes that can carry large amounts of DNA for gene
PT therapy
XX
XX Claim 8; Fig 6; 540pp; English.
XX
XX The present sequence represents the HC-contig derived from normal human
CC chromosome 10, 10q25.2 region. This region can be naturally mutated to
CC produce an unusual chromosomal marker referred to as mardel(10). The
CC mardel(10) marker is mitotically stable and contains a functional
CC neocentromere at a location regarded as non-centromeric. This
CC neocentromere maps to q25.2 on chromosome 10. The specification describes
CC nucleic acid sequences derived from a eukaryotic chromosome, including a
CC neocentromere or its functional derivative or hybrid, that are able, in
CC a compatible cell, of replicating, acting as extra-chromosomal element
CC and segregating during cell division. The sequences can be used to
CC construct artificial chromosomes for use in gene therapy comprising a
CC replicable, segregating nucleic acid that confers a specific phenotype
CC on cells. Human artificial chromosomes can propagate in human cells and
CC carry large amounts of DNA (e.g. therapeutic genes), and, being
CC extra-chromosomal, they are not mutagenic. The artificial chromosomes
CC are also useful for generation of transgenic plants and animals, in
CC production of proteins and to make diagnostic reagents, e.g. for
CC expression of cytokines, receptors and growth factors, or to increase
CC the copy number of a gene in a cell. The constructs may also be
CC used for functional and structural analysis of chromosomes.
XX
XX Sequence 80595 BP; 23183 A; 16613 C; 16824 G; 23975 T; 0 other;
SQ
Query Match 7.8%; Score 404.6; DB 20; Length 80595;
Best Local Similarity 79.7%; Pred. No. 6.3e-62;
Matches 551; Conservative 0; Mismatches 84; Indels 56; Gaps 4;
XX 3363 AGTGTGCGCGAATTTATTTCTGTGTGGTGTCTCGCTGACTCCCAAGATGAA 3422
DB 60040 ACTGTGTCTGGAATTTGTTCTTCCCGTGGGTCTTGTGCTGACTTCAAGATGAA 59981
XX 3423 ACCGCGACCCCTTGGAGTGTCTACAGTCTTAAAGATGGTGTCTCAGAGTTTCTTC 3482
DB 59980 GCCACAGACCCCTTGGAGTGTCTACAGTCTTAAAGATGGTGTCTGAGTTTCTTC 59921

Qy 3483 CTTCAGATGTTTCAGATGTCTCCGGAGTTTCTCCCTATGTTGAGTTCGTGCTCGCTGA 3542
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 59920 CTTTCAGATATTCAGTGTGTCCAGAGTTTCTTCTCTCTGTTGGTGTCTGCTCAGTGA 59861
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3543 CTTCAACAATGAAGCCGACAGACCTTTGTCAGTGTGTGTGACAGTTCCTTAAGGAGTGC 3602
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 59860 CTTTCAGAGTGAAGCTGCAGAGCTTTCGAGTGA--GTGTTACAGCTCTTAAAGCGAGC 59803
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3603 GTCAGAGTGTGTTGTTCTCCCGTAGGTTCGTTCTCGCTGATGTCAGAAATGAAGC 3662
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 59802 ATCGGGAGTCTTGTGTTCTTCCAGTGGTTCGTTGGTCT-CTTACGTTTCAGAAATGAAGC 59744
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3663 TGCAGACCCCTCGCGTAAAGTGTACAGCTCATAAGAGTAGTGCAAAACCCAAACAGTGAAGC 3722
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 59743 CACAGACCCCTCGCAGTGTGTACAGCTCATAAGAGTAGTGACACACAAGATGAAGC 59684
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3723 AGTAGCAAGATTTATATGAAGAGCAAAAGAGCTTCCGCCACCATAGAAACAGGACC 3782
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 59683 AGCAGCAAGATTTATGTTGAAGAGCAAAAGAGCTTCCACAGTGTGGAAGGGGACC 59624
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3783 AGAATTGGTTCGCTGCTGCTG----TGGTAGCCAGCTTTTATTCCTTATTTGGCCACACC 3838
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 59623 CGAGTGGTTCGCCACTGCTGGCTCGGTGGCCAGTCTTACTCCTTATTTGACCCCGCC 59564
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3839 CACATCTGCTGATTGGCCCATTTTACAGAAATGCTGATTGGTCCATTATATAGCTGCTG 3898
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 59563 CACATCTGCTGATTGGTCCA----- 59543
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3899 ATTGGTCCGTTTTTACAGAGTGTGATTGGTGCATTTTACAAATCCTTTAGCTAGACACAGA 3958
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 59542 -----TTTTACAGAGTGTGATTGGTGCATTTTACAAATCCTTTAGCTAGACACAGA 59493
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3959 GTGCTGATTGGTGCCTTTTAAATCCTTTAGCTAGACACAAAAGTTCTACAAGTCCCCACC 4018
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 59492 GTGCTGATTGGTGCATTTTACAATCCTTTAGCTAGACACAAAAGTTCTCCAAGTCCCCACC 59433
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 4019 CAACCCAGAAAGCTCCGCTCGCTTCACTCTC 4049
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 59432 TGACCCAGAGCCAGCTGGCTTCACTCTC 59402
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Search completed: September 22, 2003, 18:13:51
Job time : 1235 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2003, 17:49:29 ; Search time 318 Seconds
(without alignments)
7241.182 Million cell updates/sec

Title: US-09-996-956-1
Perfect score: 5217
Sequence: 1 gaactttaaatatccctc.....taaaaaaaaaaaaaaaaaaa 5217

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PT05_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	404.6	7.8	80246	3	US-09-078-294-4
C 2	404.6	7.8	80595	3	US-09-078-294-3
C 3	338.4	6.5	1043	4	US-09-422-576D-6
C 4	311.4	6.0	1091	4	US-09-422-576D-5
C 5	274.6	5.3	801	4	US-09-422-576D-7
C 6	270.4	5.2	2834	3	US-09-305-384-5
C 7	270.4	5.2	6235	3	US-09-305-384-5
C 8	270.4	5.2	6235	4	US-09-525-160B-6
C 9	270.4	5.2	6679	3	US-09-305-384-1
C 10	270.4	5.2	6679	4	US-09-525-160B-5
C 11	259	5.0	90541	4	US-09-759-359A-3
C 12	255	4.9	1165	4	US-09-422-576D-25
C 13	255	4.9	1831	4	US-09-422-576D-1
C 14	205.6	3.9	2381	2	US-08-736-770-4
C 15	176.2	3.4	19650	4	US-09-819-989-3
C 16	174	3.3	46718	4	US-09-816-093-3
C 17	172.8	3.3	9365	4	US-09-608-285A-8
C 18	172.8	3.3	9365	4	US-09-350-836B-8
C 19	172.8	3.3	9365	4	US-09-370-265-8
C 20	172.8	3.3	9365	4	US-09-557-800C-8
C 21	172.8	3.3	14747	4	US-09-608-285A-42
C 22	172.8	3.3	14747	4	US-09-557-800C-42
C 23	172.8	3.3	15977	4	US-09-608-285A-59
C 24	172.6	3.3	35060	3	US-08-814-095-7
C 25	170.8	3.3	43950	4	US-09-735-934A-3
C 26	170.8	3.3	43950	4	US-10-060-332-3
C 27	170.2	3.3	84495	4	US-09-797-906-3

C 28	169.8	3.3	7705	2	US-08-687-080-115	Sequence 115, Appl
C 29	169.6	3.3	29629	4	US-09-729-995-3	Sequence 3, Appli
C 30	169.4	3.2	13158	4	US-08-687-080-105	Sequence 105, App
C 31	169	3.2	16063	4	US-09-801-052-3	Sequence 3, Appli
C 32	168.8	3.2	9573	4	US-09-220-132-168	Sequence 168, App
C 33	168.2	3.2	21234	4	US-09-810-671-3	Sequence 3, Appli
C 34	168.2	3.2	29629	4	US-09-729-995-3	Sequence 3, Appli
C 35	167.8	3.2	4129	2	US-08-370-319C-12	Sequence 12, Appli
C 36	167.8	3.2	4129	3	US-09-224-834-12	Sequence 12, Appli
C 37	167.6	3.2	39882	4	US-09-820-924-3	Sequence 3, Appli
C 38	167	3.2	70000	4	US-09-851-896-3	Sequence 3, Appli
C 39	166.4	3.2	36741	3	US-09-301-665-3	Sequence 3, Appli
C 40	166.2	3.2	18000	4	US-09-657-346A-17	Sequence 17, Appli
C 41	165.8	3.2	62804	4	US-09-800-960-3	Sequence 3, Appli
C 42	165.6	3.2	162450	4	US-09-345-882-1	Sequence 1, Appli
C 43	165.4	3.2	20966	4	US-09-776-976-7	Sequence 7, Appli
C 44	165.4	3.2	20966	4	US-09-909-547-7	Sequence 7, Appli
C 45	165.4	3.2	20966	4	US-09-569-852B-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-09-078-294-4/c
; Sequence 4, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 80246
; TYPE: DNA
; ORGANISM: Nucleotide sequence of NC-contig
US-09-078-294-4

Query Match	7.8%	Score 404.6;	DB 3;	Length 80246;
Best Local Similarity	79.7%	Pred No. 1.3e-77;		
Matches	551;	Conservative	0;	Mismatches 84;
Indels	56;	Gaps	4;	
QY	3363	AGTGTGCCCGGAATTTATTTCTTCTGGGGTTCTTGTCTCGCTGCTCAAGAAATGAA	3422	
Db	59807	ACTGTGTCTGGAATTTGTTCCTTCGGTGGTCTTGTCTTCTTCAAGAAATGAA	59748	
QY	3423	ACCGCAGACCTTGAGGTGAGTGCACAGTCTTAAAGATGGTGTTCAGAGTTTGTTC	3482	
Db	59747	GCCACAGACCTTGGGTGAGTGTACAGTCTGAAGATGGTGTCTGGAGTTTGTTC	59688	
QY	3483	CTTCAGATGTTTCAGATGTCGGAGTTCCTCCCTATGCTGAGTTCGTGCTCGCTGA	3542	
Db	59687	CTTCAGATGTTTCAGTGTGTCAGAGTTTCTTCTTCTGGTGGTTCCTGCTCAGTGA	59628	
QY	3543	CTTCAACAAATGAAGCCGACACCTTTGTCAGTGTGTCAGTGTTCCTTAAAGCAGGTGC	3602	
Db	59627	CTTCAGAGGTGAAGCTGCAGAGTTCGCAGTTCGCGAGTGA--GTGTTACAGCTTTAAAGCAGCAC	59570	
QY	3603	GTCAGAGTTTGTTCCTTCCTCCGGTAGTGTGCGGTGCTCGGTGATGTCAGGAATGAAGC	3662	
Db	59569	ATCGGGAGTTCTTTCTTCCTCCAGTGGTGTGCTGTCT--GTTGATTCAGGAATGAAGC	59511	
QY	3663	TCCAGACCTCGCGTAAAGTGTACAGCTCATAAAGGTAGTGCACCAACCAACACGTGAGC	3722	
Db	59510	CACAGACCTCGCAGTGAGTGTACAGCTCATAAAGGTAGTGTGACACACAAAGATGAGC	59451	
QY	3723	AGTAGCAAGATTTATTTATGAAGAGCAAAAGCAAAAGCTTCCCCACCATAGAAAGGACC	3782	

SOFTWARE: FACEMATCH VERSION 3.1

; SOFTWARE: PASCALCII VERSION 3.1


```

RESULT 13
US-09-422-576D-1/c
; Sequence 1, Application US/09422576D
; Patent No. 6395549
; GENERAL INFORMATION:
; APPLICANT: Tuan, Dorothy
; APPLICANT: Long, Qiamong
; APPLICANT: Bengra, Chikh
; TITLE OF INVENTION: Long Terminal Repeat, E
; TITLE OF INVENTION: Recombinant Vectors
; FILE REFERENCE: M0351-205010
; CURRENT APPLICATION NUMBER: US/09/422,576D
; CURRENT FILING DATE: 1998-10-21
; PRIOR APPLICATION NUMBER: US 60/105,256
; PRIOR FILING DATE: 1998-10-22
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1831
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-422-576D-1

```

RESULT 14
 US-08-736-770-4/c
 ; Sequence 4, Application US/08736770
 ; Patent No. 5871965
 ; GENERAL INFORMATION:
 ; APPLICANT: Bandman, Olga
 ; APPLICANT: Au-Young, Janice
 ; APPLICANT: Hillman, Jennifer L.
 ; TITLE OF INVENTION: NOVEL HUMAN GUANYLATE BINDING PROTEINS
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSER: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: US
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq Version 1.5
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/736,770
 ; FILING DATE: Filed Herewith
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2003, 23:19:09 ; Search time 1780 Seconds
(without alignments)
7209.771 Million cell updates/sec

Title: US-09-996-956-1

Perfect score: 5217

Sequence: 1 gaactttaaataatccctc.....taaaaaaaaaaaaaaa 5217

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1660708 seqs, 1229959015 residues

Total number of hits satisfying chosen parameters: 3321416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA:

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5217	100.0	5217	10	US-09-996-956-1
2	4957	95.0	5092	10	US-09-996-956-3
3	4758.4	91.2	32463	13	US-09-996-956-5
4	463	8.9	2057	13	US-10-027-632-100451
5	462	8.9	250000	12	US-10-225-810-26
6	421.2	8.1	33206	10	US-09-880-107-2380
7	404.6	7.8	80246	11	US-09-728-552-4
8	404.6	7.8	80595	11	US-09-728-552-3
9	394.8	7.6	650	13	US-10-027-632-226741
10	390.6	7.5	32212	11	US-09-764-891-8770
11	386.2	7.4	6138	11	US-09-764-891-8066
12	363.6	7.0	1691139	14	US-10-067-514-1
13	360.2	6.9	3037	13	US-10-027-632-116187
14	359.8	6.9	599	9	US-09-864-761-7986
15	359.4	6.9	142299	11	US-09-911-077A-14
16	350.8	6.7	861	13	US-10-027-632-154541

17	350.8	6.7	861	13	US-10-027-632-154542	Sequence 154542,
18	350.8	6.7	861	13	US-10-027-632-154543	Sequence 154543,
19	350.8	6.7	861	13	US-10-027-632-154544	Sequence 154544,
20	350	6.7	2044	13	US-10-027-632-99305	Sequence 99305, A
21	350	6.7	2044	13	US-10-027-632-99306	Sequence 99306, A
22	350	6.7	28818	10	US-09-764-877-2266	Sequence 2266, Ap
23	334.6	6.4	687	13	US-10-027-632-205124	Sequence 205124,
24	319	6.1	829	13	US-10-027-632-153282	Sequence 153282,
25	318.8	6.1	460	9	US-09-864-761-177	Sequence 177, App
26	314	6.0	1503841	9	US-09-795-668-1	Sequence 1, Appli
27	314	6.0	1503841	9	US-09-795-668-1	Sequence 1, Appli
28	314	6.0	1503841	10	US-09-946-807-1	Sequence 38, Appl
29	312	6.0	128779	15	US-10-081-327-38	Sequence 3, Appli
30	311	6.0	378361	11	US-09-901-136-3	Sequence 170030
31	309.2	5.9	850	13	US-10-027-632-170030	Sequence 1878, Ap
32	308.2	5.9	6766	10	US-09-764-847-1878	Sequence 1878, Ap
33	308.2	5.9	6766	14	US-10-092-154-1878	Sequence 13, Appl
34	307.2	5.9	175590	11	US-09-911-077A-13	Sequence 3474, Ap
35	305.4	5.9	19172	10	US-09-764-877-3474	Sequence 6889, Ap
36	301.6	5.8	573	9	US-09-864-761-6889	Sequence 188551,
37	299.8	5.7	678	13	US-10-027-632-188551	Sequence 7928, Ap
38	296.4	5.7	500	9	US-09-864-761-7928	Sequence 13419, A
39	296.6	5.7	599	14	US-10-198-846-13419	Sequence 6749, Ap
40	296.6	5.7	864	14	US-10-198-846-6749	Sequence 3, Appli
41	294.8	5.7	10278	12	US-09-820-788-3	Sequence 200619,
42	294.6	5.6	632	13	US-10-027-632-200619	Sequence 182738,
43	293	5.6	603	13	US-10-027-632-182738	Sequence 182739,
44	293	5.6	603	13	US-10-027-632-182739	Sequence 1938, Ap
45	290.4	5.6	9180	12	US-10-311-455-1938	

ALIGNMENTS

RESULT 1

US-09-996-956-1
; Sequence 1, Application US/09996956
; Patent No. US20020155463A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc
; TITLE OF INVENTION: Prostate Polynucleotides and Uses
; FILE REFERENCE: 90 301 R1
; CURRENT APPLICATION NUMBER: US/09/996.956
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: 60/250,354
; PRIOR FILING DATE: 2001-12-01
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 5217
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-996-956-1

Query Match	100.0%	Score 5217;	DB 10;	Length 5217;
Best Local Similarity	100.0%	Pred No. 0;		
Matches 5217;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
Qy	1	GAACCTTTAAATATCCCTCAGTCTCTGTTAAATTCATGAGCAAGCACTCT	60	
Db	1	GAACCTTTAAATATCCCTCAGTCTCTGTTAAATTCATGAGCAAGCACTCT	60	
Qy	61	GGCACCAGTTTGGAACTGCAAGTTTAAAGTCATAAATTCATGAAATGATAGCAA	120	
Db	61	GGCACCAGTTTGGAACTGCAAGTTTAAAGTCATAAATTCATGAAATGATAGCAA	120	
Qy	121	GGTGGAGTTTAAAGAGCTATTATAGGTCCTGGACAGCATCTTTTCAATAGGC	180	
Db	121	GGTGGAGTTTAAAGAGCTATTATAGGTCCTGGACAGCATCTTTTCAATAGGC	180	
Qy	181	AGCAACCTTTTGGCCCTATGCCGTAACCTGTGTGCAACTTCTCTAATGGGAATAG	240	
Db	181	AGCAACCTTTTGGCCCTATGCCGTAACCTGTGTGCAACTTCTCTAATGGGAATAG	240	

QY	241	TTAAGCAGATTCATAGAGCTGAATGATGATAAAATTGTACTACGAGATGCACCTGGGACTCAAC	300
Db	241		
QY	301	GTGACCTTATCAAGTGAGATGGAGTCTTGCCCTGTCTCCAAAGCTGGAGGCCAATGGTGT	360
Db	301		
QY	361	GATCTTGGCTCAGCTGCNAACCTCCACCTCCAGGTTCAAGGTTTCTCCTGCCCTCAGCCTC	420
Db	361		
QY	421	CCAAGTAACCTGGGATTACAGCAGGCTTGGTGCAATTTGACACTTCATGATATCAGCCAAAG	480
Db	421		
QY	481	TGGAACCTAAAACAGCTCCCTGGAGAGGACTATGACATCATCAGGTTGGAGTCTCCAGG	540
Db	481		
QY	541	GACAGCGGACCTTTGGAAGGACTAGAAAGTGTGAAATCTATTAGTCTTCGATATGAA	600
Db	541		
QY	601	ATTCTCTGCTCTGTAAAGCAATTCATATTTACAAGACACAGGCCCTACTCCTAGGGCAG	660
Db	601		
QY	661	CAAAAAGTGGCAACAGCAAGCAGAGGGGAAAGAGATCATGAGGCATTTTCAGAGTGCAC	720
Db	661		
QY	721	GTCTTTTCATATATTTCTCAATGCCGTATGTTTGGTTTATTTTGGCCCAAGCATACAAT	780
Db	721		
QY	781	CTGCTCAAGAAAAAAATCTGGAGAAAAACAAAGGTGCCCTTGGCCAAAGTATGTTCTTT	840
Db	781		
QY	841	TTTGACAAGCCCTGAGATTTCTGAGGGGAATTCACATAAATGGGATCAGGTCAATTCATTT	900
Db	841		
QY	901	ACGTTGTGCAAAATATGATTTAAAGATACAACCTTTGCAGAGAGCATGCTTTCCTAAGG	960
Db	901		
QY	961	GTAGGCACGTGGAGGACTAAGGGTAAAGCATTTCTCAAGATCAGTTAAATCAAGAAAGGTG	1020
Db	961		
QY	1021	CTCTTTGCATCTGAAATGCCCTTGTGCAAAATTTGGTTATATTGATTAATTTACACT	1080
Db	1021		
QY	1081	TAATGGAACAACCTTTAACTTACAGATGAACAAACCCACAAAGCAAAAAATCAAAAGC	1140
Db	1081		
QY	1141	CCTACCTATGATTTTCATATTTTCTGTAACTGGATTAAAGGATTCCTGCTCTTTTGG	1200
Db	1141		
QY	1201	GCATAAATGATATGAATATTTCCAGGTATTTGTTAAATGAGGCCCATCTACAAAT	1260
Db	1201		
QY	1261	CTTAGCAATCTTTGGATATTTCTAAAATTCAGCTGGACATTCGCTAAATGTTTTTATA	1320
Db	1261		
QY	1321	TACATCTTTGCTAGAAATTTCAAATTTTAACTATGTGAATTTAGTAAATTTAGCTGTGCTGA	1380
Db	1321		
QY	1381	TCAATTCAAAAGCAATTAATTTCTTAAATTTTAACTATGTGAATTTAGTAAATTTAGCTGTGCTGA	1440
Db	1381		
QY	1441	ATATCTACACATACAATTAATTTAGTAAATTTTAACTATGTGAATTTAGTAAATTTAGCTGTGCTGA	1500
Db	1441		
QY	1501	GTCTTTGTGATTTGTTTGAAGAACTGAGAGTTTAAATTTCAATATTTACTTTGATCAAAAAA	1560
Db	1501		
QY	1561	TTGTGGGAACAATCCAGCATTAATTTGATGTGATTTGTTTATGTACATAGGAGTCTTA	1620
Db	1561		
QY	1621	AGCTTTGTGCTTTGAAAGTCTTTTGTACTTGTAGTCCCATGTTTAAATTTACTTATATATC	1680
Db	1621		
QY	1681	TAAAGCATTTTATGTTTTCATTAATTTACATGATGCTAAATTTATGCAATTTATAACAAA	1740
Db	1681		
QY	1741	TATTAAGATTTTTCGAATAGAATTTGTAATTTGTCACATACATAGAAATGAAAGTTCA	1800
Db	1741		
QY	1801	TTTCGTTAAAGCAAGATGCTGGGTGAAGAGTCTTTTGTGTTGAAAGATCACTAGATAGT	1860
Db	1801		
QY	1861	AGAGGCAAGACTTCTAGTCCCTAACTACCTTTAATAGCCATGCTGCTCAGGTGAAGTC	1920
Db	1861		
QY	1921	AGTGAACCCATCTCATCTCCTCATACTTTTTCATCTCTAAAATGAGGATATAATTTAA	1980
Db	1921		
QY	1981	GCCTTTCAATTTTATTTTGGATAGAGTTTGTCTTGTACCCAGGTGGAGTG	2040
Db	1981		
QY	2041	CAATGGCAGCATCTAGCTCAGTGCACCCCTCTGCTTCCCTCGGTTCAAGTATCTCCTG	2100
Db	2041		
QY	2101	CTTCAGCTCCCAAGTAGCCGGATTTACAGGTGCCGCCACCATCTGCTAATTTT	2160
Db	2101		
QY	2161	GTATTTTCAACATGTTGGCAGGCTGGTCTCGAACCCCTACCTCAGGTGATCCCTCGCCT	2220
Db	2161		
QY	2221	CGCCCTCTCAAAGTCTGGGATTTACAGGTGTAGCCACCCAGCCAGCCCAATATCAGTT	2280
Db	2221		
QY	2281	TTTCTTTTAAACAGGCTAACAAATCAAAATCTAGCTAGGGAGGAGAAAAAAT	2340
Db	2281		
QY	2341	AAGGCACTGTTTATGTAAACAGGCTCTTGTGCAATCTAGTGGCGAGACAAATTAACAGTA	2400
Db	2341		
QY	2401	AGAAATCAATCTTTTCATATATCTCTTGTGAGAAATACATAAAATCCCAAAATGGCTAT	2460
Db	2401		

[illegible]

Qy	4621	AAAGAGTTTGTCAATATAAGATTCCTTCTTTGTTAAACATCTACCTTGGGGCTTCATTT	4681
Db	4621	AAAGAGTTTGTCAATATAAGATTCCTTCTTTGTTAAACATCTACCTTGGGGCTTCATTT	4680
Qy	4681	ATAATCAAAAGGAGTACTGTAACCTGTCAAAAAAAGCTACCTGTGACAATATATATGT	4740
Db	4681	ATAATCAAAAGGAGTACTGTAACCTGTCAAAAAAAGCTACCTGTGACAATATATATGT	4740
Qy	4741	GATGGTTACCTGCAGTAGAGTGGTGGCAATAAATAAATATATCACGAATGAACCCG	4800
Db	4741	GATGGTTACCTGCAGTAGAGTGGTGGCAATAAATAAATATATCACGAATGAACCCG	4800
Qy	4801	AGCAGAACTGTCAAGAGAAATGGTCAGAAATTCACACTCTGAAGAACACGGCTATACAGTAA	4860
Db	4801	AGCAGAACTGTCAAGAGAAATGGTCAGAAATTCACACTCTGAAGAACACGGCTATACAGTAA	4860
Qy	4861	TAATCATATAAATAGCCACTCAATCCAAAACATCACTGGGGGACATTGTACATATATAA	4920
Db	4861	TAATCATATAAATAGCCACTCAATCCAAAACATCACTGGGGGACATTGTACATATATAA	4920
Qy	4921	TCAGTGGAGATGTGATTGAAGCACACAGGCTTAAGTGAATCTCTAGAGAGCTAATTGATTC	4980
Db	4921	TCAGTGGAGATGTGATTGAAGCACACAGGCTTAAGTGAATCTCTAGAGAGCTAATTGATTC	4980
Qy	4981	ATTTTTATGAAATTTTACTATTATTTAAATGTCAATCCCTGACCATCTTGAACCTTTACTT	5040
Db	4981	ATTTTTATGAAATTTTACTATTATTTAAATGTCAATCCCTGACCATCTTGAACCTTTACTT	5040
Qy	5041	GAAGATTATTTTTTTTTTTTAAATCACTGTTTATAGATTATAGGTATCTGGCTTTGTT	5100
Db	5041	GAAGATTATTTTTTTTTTTTAAATCACTGTTTATAGATTATAGGTATCTGGCTTTGTT	5100
Qy	5101	TTTCTTTTTTATCTATGTATGATTTTTATTTTTTTATGCACTGCTCTTAAGCTTCATCAA	5160
Db	5101	TTTCTTTTTTATCTATGTATGATTTTTATTTTTTTATGCACTGCTCTTAAGCTTCATCAA	5160
Qy	5161	TGAGAGAAATGTATTAATAATCCATTTATTTCTTACCCCTAAAAAATAAAAAA	5217
Db	5161	TGAGAGAAATGTATTAATAATCCATTTATTTCTTACCCCTAAAAAATAAAAAA	5217

RESULT 2

RESOLUT 2
US-09-996-956-3

03-03-330-330-3
: Sequence 3 Application IIS/09996956

; sequence 3, Application US
: Patent No. US20020155463A1

FILE NO. 032002013
: GENERAL INFORMATION:

APPLICANT: Origene Technologies, Inc.

APPLICANT: Origene Technologies, Inc

FILE OF INVENTION: PLUS
FILE REFERENCE: 9U 301 P1

FILE REFERENCE: 90 301 RI
CURRENT APPLICATION NUMBER: US/09/0906 956

1. CURRENT FILING DATE: _____

: CURRENT FILING DATE: 2001-11-30
 : PRIOR APPLICATION NUMBER: 60/250 364

; PRIOR APPLICATION NUMBER: 60/2
 : PRIOR FILING DATE: 2001-12-01

; PRIOR FILING DATE: 2001-11-15
; NUMBER OF SEQ ID NOS: 15

; NUMBER OF SEQ ID NOS: 16

SOFTWARE: F

; SEQ ID NO 3

```

; LENGTH: 50

```

```

; TYPE: DNA

```

; ORGANISM: HO

Query Match	95.0%	Score 4957;	DB 10;	Length 5092;
Best Local Similarity	97.6%	Pred. No. 0;		
Matches 5092;	Conservative	0;	Mismatches	0;
Indels 125;	Gaps	1;		

[illegible]

QY I GAAACTTTAAAAATATCCCTCAGTGTCCCTGTTAAATCATGGTAGTGCCCCAAGGCACTCT 60

.....

QY 61 GGCACCCAGTTTGGAACTGCAGTTTAAAGTCATAAATTGAATGAAATGATAGCAA 120

QY 423 AAGTAACTGGGATTACAGCAGGCTTGGTGCAATTTGACACTTCATGATATACGCCAAAGTG 482
| | | | |
DB 27688 AAATATTCTCACCCAGCAGCCTTGGTGCAATTTGACACTTCATGATATACGCCAAAGTG 27747
QY 483 GAATTAACACAGCTCCTGGAGAGGACTATGACATCATCAGGTGGGAGTCTCCAGGGA 542
| | | | |
DB 27748 GAATTAACACAGCTCCTGGAGAGGACTATGACATCATCAGGTGGGAGTCTCCAGGGA 27807
QY 543 CAGCGGCCCTTGGAAAGGACTAGAAAGTGTGAATCTAATTAGTCTTCATATGAAT 602
| | | | |
DB 27808 CAGCGGCCCTTGGAAAGGACTAGAAAGTGTGAATCTAATTAGTCTTCATATGAAT 27867
QY 603 TCTCTGCTCTGTAAAAGCAATTCATATTTACAAGACACAGCGCTACTCTCTAGGCGCAGCA 662
| | | | |
DB 27868 TCTCTGCTCTGTAAAAGCAATTCATATTTACAAGACACAGCGCTACTCTCTAGGCGCAGCA 27927
QY 663 AAAAGTGGCAACAGCGCAAGAGAGGGAAGAAAGAGATCATGAGGCAATTCAGAGTGCATGT 722
| | | | |
DB 27928 AAAAGTGGCAACAGCGCAAGAGAGGGAAGAAAGAGATCATGAGGCAATTCAGAGTGCATGT 27987
QY 723 CTTTTCATATATTTCTCAATGCCGTATGCTTTTGGTTTATTTTGGCCAGCATATAACAATCT 782
| | | | |
DB 27988 CTTTTCATATATTTCTCAATGCCGTATGCTTTTGGTTTATTTTGGCCAGCATATAACAATCT 28047
QY 783 GCTCAAGAAAAAATCTGGAGAAAAACAAAGGTGCCCTTTGCCAATGTTATGTTTCTTTT 842
| | | | |
DB 28048 GCTCAAGAAAAAATCTGGAGAAAAACAAAGGTGCCCTTTGCCAATGTTATGTTTCTTTT 28107
QY 843 TGACAGCCCTGAGATTTCTGAGGGGAATTCACATAAATGGGATCAGGTCAATTCATTTAC 902
| | | | |
DB 28108 TGACAGCCCTGAGATTTCTGAGGGGAATTCACATAAATGGGATCAGGTCAATTCATTTAC 28167
QY 903 GTTGTGCAAAATATGATTTAAAGATACAACCTTTGCAGAGAGCATGCTTCCCTAAAGGT 962
| | | | |
DB 28168 GTTGTGCAAAATATGATTTAAAGATACAACCTTTGCAGAGAGCATGCTTCCCTAAAGGT 28227
QY 963 AGGCACGTGGAGGACTAAGGTAAGCAATCTTCAAGATCAGTTAATCAAGAAAGGTGCT 1022
| | | | |
DB 28228 AGGCACGTGGAGGACTAAGGTAAGCAATCTTCAAGATCAGTTAATCAAGAAAGGTGCT 28287
QY 1023 CTTTGCATTCGAATGCCCTTGTGCAAAATTTGGTTATATTTGATTAATTAATTTACACTTA 1082
| | | | |
DB 28288 CTTTGCATTCGAATGCCCTTGTGCAAAATTTGGTTATATTTGATTAATTAATTTACACTTA 28347
QY 1083 ATGGAACAACCTTTAACTTACAGATGAACAAACCCACAAAGCAAAAATCAAAAGCCC 1142
| | | | |
DB 28348 ATGGAACAACCTTTAACTTACAGATGAACAAACCCACAAAGCAAAAATCAAAAGCCC 28407
QY 1143 TACCTATGATTTTCATATTTCTGRTAACTGGATTAAGGATTCCTGCTGCTTTTGGGC 1202
| | | | |
DB 28408 TACCTATGATTTTCATATTTCTGRTAACTGGATTAAGGATTCCTGCTGCTTTTGGGC 28467
QY 1203 ATAAATGATAATGAATATTTCCAGGTATTTCTTAAATGAGGCCCATCTACAAATCT 1262
| | | | |
DB 28468 ATAAATGATAATGAATATTTCCAGGTATTTCTTAAATGAGGCCCATCTACAAATCT 28527
QY 1263 TAGCAATACTTTGGATTAATTTCAAAATTCAGCTGGACATTCGCTAAATTTGTTTTATATA 1322
| | | | |
DB 28528 TAGCAATACTTTGGATTAATTTCAAAATTCAGCTGGACATTCGCTAAATTTGTTTTATATA 28587
QY 1323 CATCTTGTCTAGAAATTTCAAAATTTAAGTATGTAATGTAATTAATGCTGTGCTGATC 1382
| | | | |
DB 28588 CATCTTGTCTAGAAATTTCAAAATTTAAGTATGTAATTAATTAATGCTGTGCTGATC 28647
QY 1383 AATTCAAAACATTTACTTTTCCCTAAATTTAGCATATGAAGTCAATAAATCAACAATAT 1442
| | | | |
DB 28648 AATTCAAAACATTTACTTTTCCCTAAATTTAGCATATGAAGTCAATAAATCAACAATAT 28707
QY 1443 ATCTACACATACAATATATAGATTTGTTTTTCATTTAATGCTTCTCATCTTAACAGAAATGT 1502
| | | | |
DB 28708 ATCTACACATACAATATATAGATTTGTTTTTCATTTAATGCTTCTCATCTTAACAGAAATGT 28767

QY 1503 CTTTGTGATTTGTTTTAGAAAACTGAGAGTTTAAATTCATAATTTACTTGATCAAAAAAT 1562
| | | | |
DB 28768 CTTTGTGATTTGTTTTAGAAAACTGAGAGTTTAAATTCATAATTTACTTTGATCAAAAAAT 28827
QY 1563 GTGGACAACATCCAGCATTAATGTATGTGATTTTATGTACATAAGAGTCTTTAAG 1622
| | | | |
DB 28828 GTGGACAACATCCAGCATTAATGTATGTGATTTTATGTACATAAGAGTCTTTAAG 28887
QY 1623 CTTGCGCTTGAAGTCTTTTGTACTTAGTCCCATGTTTAAATTTACTTACTTTATATCTA 1682
| | | | |
DB 28888 CTTGCGCTTGAAGTCTTTTGTACTTAGTCCCATGTTTAAATTTACTTACTTTATATCTA 28947
QY 1683 AAGCATTTATGTTTTTCAATTTACATGATGCTTAATTTATGSCAATTAACAATA 1742
| | | | |
DB 28948 AAGCATTTATGTTTTTCAATTTACATGATGCTTAATTTATGSCAATTAACAATA 29007
QY 1743 TTAAGATTTTCAAAATAGAATATGTGAATTTCTCACATACATAGAAATGAAAAGTTCAAT 1802
| | | | |
DB 29008 TTAAGATTTTCAAAATAGAATATGTGAATTTCTCACATACATAGAAATGAAAAGTTCAAT 29067
QY 1803 TCGTAAAGCAAGATGCTGGGTGAAAGAGTCTTTTGTATTGAAGATCCTAGATTTAGTAG 1862
| | | | |
DB 29068 TCGTAAAGCAAGATGCTGGGTGAAAGAGTCTTTTGTATTGAAGATCCTAGATTTAGTAG 29127
QY 1863 AGGCAAGACCTCTAGTCCCTTAATCTACCCCTTAATAGCCATGTGGTCAAGTCAAGTCAG 1922
| | | | |
DB 29128 AGGCAAGACCTCTAGTCCCTTAATCTACCCCTTAATAGCCATGTGGTCAAGTCAAGTCAG 29187
QY 1923 TGAACCCATCTCAATTTCTCTCATACTTTTTTTCATCTCTAAAATGAGGATATAATTTAAGC 1982
| | | | |
DB 29188 TGAACCCATCTCAATTTCTCTCATACTTTTTTTCATCTCTAAAATGAGGATATAATTTAAGC 29247
QY 1983 TCTTCATTTTTTTTTTTTTTTTGGATAGAGTTTTTGTCTCTCCACCCAGGTGGAGTGCA 2042
| | | | |
DB 29248 TCTTCATTTTTTTTTTTTTTTTGGATAGAGTTTTTGTCTCTCCACCCAGGTGGAGTGCA 29307
QY 2043 ATGCGACGATCTCAGCTCAGTCAACCCCTCTGCTTCCTCGGTTCAAGTGATTCCTCGCT 2102
| | | | |
DB 29308 ATGCGACGATCTCAGCTCAGTCAACCCCTCTGCTTCCTCGGTTCAAGTGATTCCTCGCT 29367
QY 2103 TCAGCCTCCCAAGTAGCGGGATTTACAGGTGCCCGCCACACACATCTGCCTAATTTTTTGT 2162
| | | | |
DB 29368 TCAGCCTCCCAAGTAGCGGGATTTACAGGTGCCCGCCACACACATCTGGCTAATTTTTTGT 29427
QY 2163 ATTTTCAACATTTTGGCAGGCTGCTCTCGAAACCCCTACCTCAGGTGATCCTCGCCTCG 2222
| | | | |
DB 29428 ATTTTCAACATTTTGGCAGGCTGCTCTCGAAACCCCTACCTCAGGTGATCCTCGCCTCG 29487
QY 2223 GCCTCTCAAAAGTGTGGGATTTACAGGTGTGAGCCACACAGCCCGCCCAATATCAGTTTT 2282
| | | | |
DB 29488 GCCTCTCAAAAGTGTGGGATTTACAGGTGTGAGCCACACAGCCCGCCCAATATCAGTTTT 29547
QY 2283 TCTTTTTTAAACAAAGGCTTAACAAATCTAGCTAGGGGAGAAAAAATAA 2342
| | | | |
DB 29548 TCTTTTTTAAACAAAGGCTTAACAAATCTAGCTAGGGGAGAAAAAATAA 29607
QY 2343 GGCACCTGTTTATGTAAACAGGCTCTGTTTGCATCTACTGGCGACACATAAAGCAGTAAG 2402
| | | | |
DB 29608 GGCACCTGTTTATGTAAACAGGCTCTGTTTGCATCTACTGGCGACACATAAAGCAGTAAG 29667
QY 2403 AATCAATCTTTTCATATATCTTCTGCGAAGATACATAAATCCCAAAATGGCTATCT 2462
| | | | |
DB 29668 AATCAATCTTTTCATATATCTTCTGCGAAGATACATAAATCCCAAAATGGCTATCT 29727
QY 2463 TCCTTTTTATCATATTTGGAGAAATTTAGCTAAAGTACAGATATTTTCTTTGGGTGATA 2522
| | | | |
DB 29728 TCCTTTTTATCATATTTGGAGAAATTTAGCTAAAGTACAGATATTTTCTTTGGGTGATA 29787
QY 2523 GACCACAAGGACCTGTTTGTGATGTTTGGCATAAAAATATACCTTTAGTTTTTACTTTT 2582
| | | | |
DB 29788 GACCACAAGGACCTGTTTGTGATGTTTGGCATAAAAATATACCTTTAGTTTTTACTTT 29847
QY 2583 GTATGTTACATGTTTAGATTTAGAAATTTAGTAGGGAGGATTTAATTAACAAAGAAC 2642

Db	29848	GTATGTTACATGTTAGATTAGAGTATGAAAATAGTAGGGAGGATTATTAAACAAGAAC	29907	Db	30928	TGCAGACCCGCGGTAAAGTCTACAGCTATAAGGTTAGTGCATAAACCCCAACACAGTGAGC	30987
Qy	2643	AGGCAAGAGAGTAGAATTAACACTCTCTTAATACCTGTGCACAAAGTAGGCTTTTCAGA	2702	Qy	3723	AGTAGCAAGATTTTATTTATGAGAGCAAAAGAACAAAGCTTCCCCACCATAGAAACGGACC	3782
Db	29908	AGGCAAGAGAGTAGAATTAACACTCTCTTAATACCTGTGCACAAAGTAGGCTTTTCAGA	29967	Db	30988	AGTAGCAAGATTTTATTTATGAGAGCAAAAGAACAAAGCTTCCCCACCATAGAAACGGACC	31047
Qy	2703	AACTCTACACCCCTACATAAACTGGATAGTTAGAAAAGACACACACCTCCCAAGGAAGCGGTT	2762	Qy	3783	AGAAATGGTGTGCTGCTGCTGTGTGAGCAGCTTTTATTTCCCTTATTTTGGCCACACCCACA	3842
Db	29968	AACTCTACACCCCTACATAAACTGGATAGTTAGAAAAGACACACCTCCCAAGGAAGCGGTT	30027	Db	31048	AGAAATGGTGTGCTGCTGCTGTGTGAGCAGCTTTTATTTCCCTTATTTTGGCCACACCCACA	31107
Qy	2763	ATGTTTTGCAGTTTGAATCAGAAGAAATAGAGCTATAGCAATCTTTCATTTCTATAGTAACAT	2822	Qy	3843	TCCTGCTGATTTGGCCCAATTTTACAGAAATGCTGATTTGGTCCATTTTATAGCGTGTGATG	3902
Db	30028	ATGTTTTGCAGTTTGAATCAGAAGAAATAGAGCTATAGCAATCTTTCATTTCTATAGTAACAT	30087	Db	31108	TCCTGCTGATTTGGCCCAATTTTACAGAAATGCTGATTTGGTCCATTTTATAGCGTGTGATG	31167
Qy	2823	TAAAGAGCCGGTTTATTTATATAGCAGTCATTAAGATTTTAAATTTTACATCTTGGCGTT	2882	Qy	3903	GTGCGTTTTCACAGAGTGCTGATTTGGTGCAATTTTACAAATCTTTAGCTTAGACACAGAGTGC	3962
Db	30088	TAAAGAGCCGGTTTATTTATATAGCAGTCATTAAGATTTTAAATTTTACATCTTGGCGTT	30147	Db	31168	GTGCGTTTTCACAGAGTGCTGATTTGGTGCAATTTTACAAATCTTTAGCTTAGACACAGAGTGC	31227
Qy	2883	CTTCTTACTCACAGATTTTCGAGAGGTAATGTAATGATCCAGAGGTGAGAATCACCTGCC	2942	Qy	3963	TGATTGCTGCTTTTATTAATCTTTTAGCTAGACACAAAGTTCTACAAGTCCCCACCCAAAC	4022
Db	30148	CTTCTTACTCACAGATTTTCGAGAGGTAATGTAATGATCCAGAGGTGAGAATCACCTGCC	30207	Db	31228	TGATTGCTGCTTTTATTAATCTTTTAGCTAGACACAAAGTTCTACAAGTCCCCACCCAAAC	31287
Qy	2943	TTTTATAATCGAATTAATTCGATGAACAAAGTTTCCACAAATTAACAGTAATAAAGA	3002	Qy	4023	CCAGAACTCCGCTGGCTTCACCTCTCGTAAGAAATTTAGGTTTCAACAAAGTTTCAAG	4082
Db	30208	TTTTATAATCGAATTAATTCGATGAACAAAGTTTCCACAAATTAACAGTAATAAAGA	30267	Db	31288	CCAGAACTCCGCTGGCTTCACCTCTCGTAAGAAATTTAGGTTTCAACAAAGTTTCAAG	31347
Qy	3003	AACATGATTTAGCACTTAATAAGCCAGGGCTGGACAGCTGTGTTACATCTTCAATC	3062	Qy	4083	TGCTAAAACCTACAGTTTCTCATCTCTGCAACTGGATTTCCACTCATGTGTTGAATCCC	4142
Db	30268	AACATGATTTAGCACTTAATAAGCCAGGGCTGGACAGCTGTGTTACATCTTCAATC	30327	Db	31348	TGCTAAAACCTACAGTTTCTCATCTCTGCAACTGGATTTCCACTCATGTGTTGAATCCC	31407
Qy	3063	CATGAACCTGTAAGTAACTGTAATCTCTATTTGGACATGTGAGGAAACCAATGGAGT	3122	Qy	4143	AGGCTCTAAGACTTAACCTTGCCATTTCTGACTTTATGTTCTCTGCAATTTTACACAAAGCT	4202
Db	30328	CATGAACCTGTAAGTAACTGTAATCTCTATTTGGACATGTGAGGAAACCAATGGAGT	30387	Db	31408	AGGCTCTAAGACTTAACCTTGCCATTTCTGACTTTATGTTCTCTGCAATTTTACACAAAGCT	31467
Qy	3123	TGATAAACAGTAGAGTTTAAATTTACTCTTCATATATTTATTTGCCCTCAATCTCACAGAC	3182	Qy	4203	ACTATCTGTCCATCTCTGTGTGTTAACTTCAGACTTAAACTTCTTTTGTATTCAATATGAC	4262
Db	30388	TGATAAACAGTAGAGTTTAAATTTACTCTTCATATATTTATTTGCCCTCAATCTCACAGAC	30447	Db	31468	ACTATCTGTCCATCTCTGTGTGTTAACTTCAGACTTAAACTTCTTTTGTATTCAATATGAC	31527
Qy	3183	ATCTCTGCTACCAAAAGCTATCATCTAGATATGCGGCATAGGATGACCTTGGGGCAC	3242	Qy	4263	CACACACTTTTGGTGTGAGGTTTGTCTATCGGTTTATTTGCTGTTTAAATAGAGAGCTTC	4322
Db	30448	ATCTCTGCTACCAAAAGCTATCATCTAGATATGCGGCATAGGATGACCTTGGGGCAC	30507	Db	31528	CACACACTTTTGGTGTGAGGTTTGTCTATCGGTTTATTTGCTGTTTAAATAGAGAGCTTC	31587
Qy	3243	ACTAGAAATCTTTGAGAGAAATCTGGCAGAGAAACAAATATTTATTCCTTCAATATAAAC	3302	Qy	4323	TTCCAGAAATTTGAGTAGATGGAAGAGAGAAAGTAGACATCTTCCCTTAAATAATTTACCATGCCT	4382
Db	30508	ACTAGAAATCTTTGAGAGAAATCTGGCAGAGAAACAAATATTTATTCCTTCAATATAAAC	30567	Db	31588	TTCCAGAAATTTGAGTAGATGGAAGAGAGAAAGTAGACATCTTCCCTTAAATAATTTACCATGCCT	31647
Qy	3303	CCAGCAATTTTACAGGTTTATTTTAACTATGAAGTATTTGTTATCTGTATCTTTTCATATA	3362	Qy	4383	TTCAAGTTCACAGCATCCCTTATCACATGGCTGTCAAGGGTGGCTCAGAAATAGGTAGAGTT	4442
Db	30568	CCAGCAATTTTACAGGTTTATTTTAACTATGAAGTATTTGTTATCTGTATCTTTTCATATA	30627	Db	31648	TTCAAGTTCACAGCATCCCTTATCACATGGCTGTCAAGGGTGGCTCAGAAATAGGTAGAGTT	31707
Qy	3363	AGTGTGCGCGGAATTTATTTCTGTTGGTGGTCTTGGTCTCGCTGACCTCCAAGATGAA	3422	Qy	4443	AAGAATTTAAAGTAAATTTGGTGAAGGATGAAAGCTTTCATCTTAAAGCTTATATTACAT	4502
Db	30628	AGTGTGCGCGGAATTTATTTCTGTTGGTGGTCTTGGTCTCGCTGACCTCCAAGATGAA	30687	Db	31708	AAGAATTTAAAGTAAATTTGGTGAAGGATGAAAGCTTTCATCTTAAAGCTTATATTACAT	31767
Qy	3423	ACCGCAGACCCCTTGAGTGAGTGTCACAGTTCTTAAAGATGGTGTGTTCAGAGTTTGTTC	3482	Qy	4503	CAACTGAAATGTAATAATTTGGAACATTTTCCAGGCATCCCTGTTTATTTTATTTGTCTCT	4562
Db	30688	ACCGCAGACCCCTTGAGTGAGTGTCACAGTTCTTAAAGATGGTGTGTTCAGAGTTTGTTC	30747	Db	31768	CAACTGAAATGTAATAATTTGGAACATTTTCCAGGCATCCCTGTTTATTTTATTTGTCTCT	31827
Qy	3483	CTTCAGATGTTTACAGTGTGTCGGAGTTTCTCCCTTATGGTGTGAGTTCGTGGTCTCGCTGA	3542	Qy	4563	CTTTCCTTGTGCTGCCCTTACCTTCAAAAGTTCATATGGCATGGTGAGTGTAGAACTGTCTGCCAA	4622
Db	30748	CTTCAGATGTTTACAGTGTGTCGGAGTTTCTCCCTTATGGTGTGAGTTCGTGGTCTCGCTGA	30807	Db	31828	CTTTCCTTGTGCTGCCCTTACCTTCAAAAGTTCATATGGCATGGTGAGTGTAGAACTGTCTGCCAA	31887
Qy	3543	CTTCAACAATGAAGCCGACACCTTTGACGTGAGTGTGTGACAGTTTCTTAAAGGCAAGTGC	3602	Qy	4623	AGAGTTTGTCAATATAGATTCCTTCTTGTAAACATTTCTACTCTTGGGGCTTCAATTTAT	4682
Db	30808	CTTCAACAATGAAGCCGACACCTTTGACGTGAGTGTGTGACAGTTTCTTAAAGGCAAGTGC	30867	Db	31888	AGAGTTTGTCAATATAGATTCCTTCTTGTAAACATTTCTACTCTTGGGGCTTCAATTTAT	31947
Qy	3603	GTCCAGAGTTGTTTGTCTCCGGTAGGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	3662	Qy	4683	AATCAAAAGGAGTACTGTAACTGTCAAAAAAGCTTACCTGTGACATATATATTATGTGA	4742
Db	30868	GTCCAGAGTTGTTTGTCTCCGGTAGGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	30927	Db	31948	AATCAAAAGGAGTACTGTAACTGTCAAAAAAGCTTACCTGTGACATATATATTATGTGA	32007
Qy	3663	TGCAGACCCCTCGGGTAAAGTGTACAGCTCATAAAGGTAGTGCAAAACCCCAACAGTGAGC	3722	Qy	4743	TGTTTACCTGTCAGTAAAGTGGTGGGCAATTAATAAATAAATAATCAACAGATGAACCGAG	4802
				Db	32008	TGTTTACCTGTCAGTAAAGTGGTGGGCAATTAATAAATAAATAATCAACAGATGAACCGAG	32067

QY	4803	CAGAACTGTCAGAGAAATGGTCAGAAATTCACACTCTGAAGAACACGGGTATACAGTAATA	4862
Db	32068	CAGAACTGTCAGAGAAATGGTCAGAAATTCACACTCTGAAGAACACGGGTATACAGTAATA	32127
QY	4863	ATCATATAAATAGCCACTCAATCCAAAACATCACTGGCGGCACTTGTACACATATATAATC	4922
Db	32128	ATCATATAAATAGCCACTCAATCCAAAACATCACTGGCGGCACTTGTACACATATATAATC	32187
QY	4923	AGTGGAGATGTGATTGAAGCACAAGCGCTTAAGTGAATGCTAGAGAGCTAAATTGATTCAT	4982
Db	32188	AGTGGAGATGTGATTGAAGCACAAGCGCTTAAGTGAATGCTAGAGAGCTAAATTGATTCAT	32247
QY	4983	TTTTATGGAAATTTTACTATTATTTTAAATGTCATCCCTGACCATCTTGAACCTTTTACTTGA	5042
Db	32248	TTTTATGGAAATTTTACTATTATTTTAAATGTCATCCCTGACCATCTTGAACCTTTTACTTGA	32307
QY	5043	AGATTTATTTTTTTTTTTTAAATCACTGTTTATTAGATTTAGGTATCTGGCTCTTGTTTT	5102
Db	32308	AGATTTATTTTTTTTTTTTAAATCACTGTTTATTAGATTTAGGTATCTGGCTCTTGTTTT	32367
QY	5103	TCCTTTTATCTATGATGATTTTTATTTTTTATGCGAGTGCCTTAAAGCTTCATCAATG	5162
Db	32368	TCCTTTTATCTATGATGATTTTTATTTTTTATGCGAGTGCCTTAAAGCTTCATCAATG	32427
QY	5163	AGAAGAAATGTAATAAATCCATTTATCTTTACCCCT	5198
Db	32428	AGAAGAAATGTAATAAATCCATTTATCTTTACCCCT	32463

RESULT 4

```

US-10-027-632-100451
; Sequence 100451, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Ma
; TITLE OF INVENTION: Polymorphisms in the
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 100451
; LENGTH: 2057
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-100451

```

	Query Match	8.9%	Score 463;	DB 13;	Length 2057;
	Best Local Similarity	80.8%;	Pred. No. 2e-83;		
	Matches 612;	Conservative 0;	Mismatches 115;	Indels 30;	Gaps 5;
QY	3358	ATATAGTGTGCCCGCAATTTATTTCTCTCGTGGGTTCTTGCTCGCTGACTCAAGA	3417		
Db	783	ATAATAATGTTCTGGAATTGGCTCTCTGTGGGTCTTGCTTGCTGACTTCAAGA	842		
QY	3418	ATGAACCCGACAGCCCTTGAGTGTAGTGTCACAGTTCCTTAAAGATGGTGTGTTCAAGTT	3477		
Db	843	ATGAACCCGACAGCCCTTGGGTGAGTGTGTAGTTCCTTAAAGATGGTGTGCCGAGTT	902		

3478	Qy	TGTTCCCTTCAGATGTTTCAGATGTCTCCGGAGTTTCTCCCTTATGGTGAGTTGCTGTCTC	3537
903	Db	TGTTCCCTTCAGATGTTTCAGATGTCTCCGGAGTTTCTCCCTTCAGTGCGGTGGTGCTCT	962
3538	Qy	GCTTGACTTCAACAATGAAGCCGAGACCTTTGCAGTGCAGTGTCTGCAGACTTCTTAAGGC	3597
963	Db	GCTGACTTCAGGAGTGAAGCCAGACACTTTCACAGTGA--GTGTCACAGCTCTTAAGGT	1020
3598	Qy	AGTGCCTGCCAGAGTTGTTGTTCCCTCCGGTAGTTCGTGGTCTCGCTCATGTGTCAGGAAT	3657
1021	Db	GGTACGTCGCGAGTGTGTTGTTCCCTCCGGTGGTTCGTGGTCTCGCTGACTTCAGGAAT	1080
3658	Qy	GAAGCTGCAGACCCCTCGCGTAAGTGTTCAGACTCATAAAGGTAGTGCAAAACCAACAG	3717
1081	Db	GAAGCTGCAGACCCCTCGAGTGGG----CAGCTCATAAAGGTAGTGCAGACCAACAG	1135
3718	Qy	TGAGCAGTAGCAAGATTTATTATGAAGAGCAAAAGACTCCGCCACCATAGAAAC	3777
1136	Db	TGAGCAGCAGCAGATTTATTGTGAAGAGTGAAGAACAAGCTTCCACAGCTGGAAG	1195
3778	Qy	GGACCAGAAATGGTGTGCTGCTGT-----GGTAGCCAGCTTTTATTCCTTATTTGGCC	3833
1196	Db	GGACCCCTAGTGGGTTGCTGCTGTGGCTTAGGTGGCCAGCTTTATTCCTTATTTGGCC	1255
3834	Qy	ACACCACATCTCTGATTGGCCCATTTTACAGAATGCTGATTGGTCCATTT-----	3886
1256	Db	CTGCCACATCTCTGCTGATTGGTGC--TCATTACAGAGTGTGATTGATGATTTACAAAC	1313
3887	Qy	-----TATAGGCTGCTGATTGGTCGGTTTTTACAGAGTGTGATTGGTGCAATTA	3936
1314	Db	TTTAGCTAGACACAGAGCGCTGATTGCTGTGTTTTTACAGAGTGTGATTGGTGCAATTA	1373
3937	Qy	CAATCTTTAGCTAGACACAGAGTGTGATTGGTGCTTTATTAATCCTTTAGCTAGACAC	3996
1374	Db	CAAACTTTAGCTAGACACAGACACTGATTTGCTGTGTTTATAACCTCTAGCTAGACAG	1433
3997	Qy	AAAAGTTCTACAAGTCCCCACCAACCCAGAAAGCTCCGCTGGCTTCACCTCTCGTAAGA	4056
1434	Db	AAAAGTCTCCAGTCCCCACCCAGCCAGAACCCAGCTGGCTTCACCTCTAAATAAGA	1493
4057	Qy	AATTGAGGTTCAACAAGTTTCAAGTGTCTAAACCTA	4093
1494	Db	TTCAATGGCTTGGGTAGTTTATTACAAACAATAATA	1530

```

RESULT 5
US-10-225-810-26/c
; Sequence 26, Application US/102225810
; Publication No. US20030157512A1
; GENERAL INFORMATION:
; APPLICANT: Birmingham, Jr., John R.
; TITLE OF INVENTION: Tramdorins and Methods of Using Tramdorin
; FILE REFERENCE: McLaugh-07165
; CURRENT APPLICATION NUMBER: US/10/225,810
; CURRENT FILING DATE: 2002-08-21
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 26
; LENGTH: 250000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (33774)..(33774)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (42953)..(43052)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45557)..(45656)

```

OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (48203)..(48302)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (49551)..(49650)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51561)..(51660)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (52722)..(52821)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (53864)..(53963)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (55290)..(55389)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (56674)..(56773)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (57879)..(57978)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (78952)..(79051)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (85316)..(85415)
OTHER INFORMATION: n is a, c, g, or t
US-10-225-810-26

Query Match 8.9%; Score 462; DB 12; Length 250000;
Best Local Similarity 84.7%; Pred. No. 3.7e-82;
Matches 574; Conservative 0; Mismatches 80; Indels 24; Gaps 4;
QY 3363 AGTGTGCCGGAATTTATTTCTCTGGTGGGTTCTTGGTCTCGTCACTCCAAAGATGAA 3422
DB 109979 AGTGTGTCGGAATGTTGTTCTTCTGGTGGGTTCTTGGTCTCGTCACTCCAAAGATGAA 109920
QY 3423 ACCGCAGACCTTGGAGTGAGTGTACAGTCTTAAAGATGGTGTTCAGAGTTGTTC 3482
DB 109919 GCCACGGCCCTCGCGTGAGTGTACAGTCTTAAAGATGGTGTTCGCGGATTTGTC 109860
QY 3483 CTTTCAGATGTTTCAGATGTGTCCGAGTTTCTCCCTTATGGTGTAGTTCGTGTCTCGCTGA 3542
DB 109859 CTTTCAGATGTTTCAGATGTGTCCGAGTTTCTTCTTGTGGGTTCTGTGTTCTGCTGA 109800
QY 3543 CTTTCACATGAAGCCGACAGACCTTTCAGTGAGTGTGTGACAGTTCCTTAAAGCGAGTGC 3602
DB 109799 CTTTCAGAGTGAAGCTTGCAGACCTTTCAGTGA--GTGTTACAGCTCTTAAAGGTGGCGC 109742
QY 3603 GTCAGAGTGTGTTGTTCTCCCGTAGGTTCTGCTGCTGTGATGTTCAGGAATGAAGC 3662
DB 109741 GTCAGAGTGTGTTGTTCTCCCGTAGGTTCTGCTGCTGTGATGTTCAGGAATGAAGC 109682
QY 3663 TGCAGACCTTCGCGGTAGTGTTCAGCTCATAAAGGTAGTGCACACCCAAACAGTGAAGC 3722
DB 109681 TGCAGACCTTTCAGCGCAATGTTTACAGCTCATACAGTGTAGTATGCGCCCAACAGAGTGAAGC 109622
QY 3723 AGTAGCAAGATTTATATGAAGGCAAAAGACAAAGCTTCCGCCCATAGAAACGGACC 3782
DB 109622 AGTAGCAAGATTTATATGAAGGCAAAAGACAAAGCTTCCGCCCATAGAAACGGACC 3782

DB 109621 AGCAGCAAGATTTATTTGTGAAGACAAAGAAAGCTTCCACAGCTTGAAGGGGACC 109562
QY 3783 AGAATTGGTTCCTGCTGCTG---TGGTAGCCACCTTTTATTCCTTATTTGGCCACACC 3838
DB 109561 CGAATGGTTCCTGCTGCTGCTCTAGTGGCCAGCTTTTATTCCTTATTTGGCCCTACC 109502
QY 3839 CACATCCTGCTGATTGGCCCA-TTTTACAGAACTGCTGATTTGGTCCATTT----- 3886
DB 109501 CACGTCCTGCTGATTGGTCCATTTTACAGAGTCTGATTTGGTGGTTCACAACTTTA 109442
QY 3887 -----TATAGCGTGTGATTGGTGGCCTTTTACAGAGTGTGATTTGGTTCATTAACAT 3940
DB 109441 GCTAAACACAGAGCGCTGATTGGTGGCTTTTACAGAGTGTGATTTGGTGGTTCACAA 109382
QY 3941 CCTTACGTAGACACAGAGTGTGATTGGTGGCCTTTTATATCCCTTTAGCTAGACACAAA 4000
DB 109381 CCTTACGTAGACACAGAGTGTGATTGGTGGCTTTTATATCCCTTTAGCTAGACAAA 109322
QY 4001 GTTCTACAAGTCCCCACC 4018
DB 109321 GTTCTCAAGTCCCCACC 109304

RESULT 6
US-09-880-107-2380/c
; Sequence 2380, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2380
; LENGTH: 33206
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 M64554
US-09-880-107-2380

Query Match 8.1%; Score 421.2; DB 10; Length 33206;
Best Local Similarity 80.3%; Pred. No. 2.3e-74;
Matches 567; Conservative 0; Mismatches 108; Indels 31; Gaps 5;
QY 3356 TCATATAAGTGTCCCGGAATTTATTTCTTCTGGTGGGTTCTTGGTCTCGCTGACTCCAA 3415
DB 1173 TCATGCAAGTGTCTCTGGAATTTGGTCTCTCAATGGTCTTGGTCTTGGTCTGACTTCAA 1114
QY 3416 GAATGAACCCGACAGACCTTGAAGTGTGATGTCTACAGTCTTAAAGATGGTGTCTCAGAG 3475
DB 1113 GAATGAAGTGTGCGGAGCTTGGCGTGAGCATTTACAGTTTAAAGATGGTGTCTCGGAG 1054
QY 3476 TTTGTTCTTCAGATGTTCAGATGTGTCCGAGCTTTCTCCCTTATGATGATTCGTGGTC 3535
DB 1053 TTTGTTCTTCAGATGTGTCCGAGCTTTCTCCCTTATGATGATTCGTGGTTCCTGGTC 995
QY 3536 TCGCTGACTTCAACATGAAGCCGACAGCTTTTGCAGTGTGTGACAGTGTCTTAAAG 3595
DB 994 TCGCTGACTTCAAGAGTGAAGCCGACAGCTTCAACAGTGA--GTGTTACAGCTCTAAAG 937
QY 3596 GCAGTCCGCTCCAGAGTTGTTGTTCTCCCGGTAGTTCGTGCTCGCTGATGTGACGA 3655
DB 936 ATGGTCCGCGAGGTTGTTCTTCTCTCTGGTGGGTTGAGGTCTTGGTGTGACTTCAGGA 877

3411	QY	TTCAAGAATGAACCGCAGACCCCTTGAGTGAGTGATCACAAGTTCTTAAAGATGGTGTTGTT	3470
5489	Db	TTCAAGAATGAACGCTG - GGACCCCTCGCAGCTGAGTGTGCAGTTCTTAAAGCTGGTGTC	5431
3471	QY	CAGAGTTGTTCCTTCAGATGTTTCAGATGTGTCGGAGTTTCTCCCTTATGTGTGAGTTGC	3530
5430	Db	TGCGGTTTGTCTTCAGATGTTTCACAGGTACCTGGAGTTTCTCTTCGTGGGGTTCG	5371
3531	QY	TGCTCTCGCTGACTTCAACAATGAAGCCGCGAGACCTTTTCGAGTGAGTGTGTGACAGTTCT	3590
5370	Db	TGCTCTCACTGACTTCAGGAGTAAGCTCGCAGACCTTCGCAGTGA - GTGTTACAGCTCT	5313
3591	QY	TAAAGCAGTGCCTCCAGAGTTGCTTTGTTCCTCCCGTAGGTTCGTGGTCTCCTGATGT	3650
5312	Db	TAAAGTGGCACCTTCGGAGTTGTTGTTCCT - CCGGGGGGTTTCGTGGTATTCTGACTT	5254
3651	QY	CAGGAATGAAGCTGCAGACCCCTCGCGTAGTGTTCACAGCTCATAAAGGTAGTGCAAACC	3710
5253	Db	GAGGAGTGAAGCTGCAGACCTTCACGCTGAGTGTTCACAGCTCATAAAGGTAGTACAGACC	5194
3711	QY	CAAAAGTGTAGCAGTAGTACGAAGATTTATTTGAAGAGCAAAAGAACAAAGCTTCCCCACCA	3770
5193	Db	CAAAAAATGAGCAACAGCAGAAGATTTACTCAAAAGAGCAAAAGAACAAAGCCTCCACAGCA	5134
3771	QY	TAGAAACGGACCAAGATTTGGTTCGCTGCTGCTG - - - TGGTAGCCAGCTTTTATTCCTTTA	3826
5133	Db	TGTAAGAGTATCCCCACGGGGTTGCCCGCTGCGCAGGTGGCAGCTTTATTCCTTTA	5074
3827	QY	TTTGGCCACACCCACATCCTGCTGATTTGGGCCCAATTTACAGAAATGCTGATGGTCCATTT	3886
5073	Db	TTTGGCCCCACCAATCCTGCTGATTTGGTCCATTTTACAGAGCGCTGATTTGGTCCATTT	5014
3887	QY	T-----ATAGCGTGCTGATTTGGTG	3905
5013	Db	TTACAGAGTGCTGATTTGGTGCTTTTACAAACCTTTTAGCTAAACACAGAGTGCTGATTTGGTG	4954
3906	QY	CGTTTTTACAGAGTGCTGATTTGGTGCAATTTACAAATCCTTTAGCTAGACACAGAGTGCTGA	3965
4953	Db	TGTTTTTACAGAGTGCTGATTTGGTGTTTGTGTTTACAAACCTTTAGCTAGACACAGAGCTGA	4894
3966	QY	TTGGTGCCCTTTAATCTCTTTAGCTAGACACAAAAGTTCTTACAAGTCCCCAAC	4018
4893	Db	TTGGTGCATTTACAATCTTTAGCTAGACAGAAAAGTTCTTCCAGGTCCCCAAC	4841

```

RESULT 11
US-09-764-891-8066/c
; Sequence 8066, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8066
; LENGTH: 6138
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-8066

Query Match          7.4%; Score 386.2; DB 11; Length 6138;

RESULT 12
US-10-067-514-1
; Sequence 1, Application US/10067514
; Publication No. US20030054531A1
; GENERAL INFORMATION:
; APPLICANT: Gretsodottir, Solveig
; APPLICANT: Jonsdottir, Sif
; APPLICANT: Reynisdottir, Sigríður Th.
; TITLE OF INVENTION: HUMAN STROKE GENE
; FILE REFERENCE: 2345,2010-003
; CURRENT APPLICATION NUMBER: US/10/067,514
; CURRENT FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: US 09/811/352
; PRIOR FILING DATE: 2001-03-19
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1691139
; TYPE: DNA
; ORGANISM: Human
US-10-067-514-1

```

```

RESULT 12
US-10-067-514-1
; Sequence 1, Application US/10067514
; Publication No. US20030054531A1
; GENERAL INFORMATION:
; APPLICANT: Gretarsdottir, Solveig
; APPLICANT: Jonsdottir, Sif
; APPLICANT: Reynisdottir, Sigridur Th.
; TITLE OF INVENTION: HUMAN STROKE GENE
; FILE REFERENCE: 2345.2010-003
; CURRENT APPLICATION NUMBER: US/10/067,514
; CURRENT FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: US 09/811/352
; PRIOR FILING DATE: 2001-03-19
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1691139
; TYPE: DNA
; ORGANISM: Human
US-10-067-514-1

```

```

RESULT 12
US-10-067-514-1
; Sequence 1, Application US/10067514
; Publication No. US20030054531A1
; GENERAL INFORMATION:
; APPLICANT: Gretarsdottir, Solveig
; APPLICANT: Jonsdottir, Sif
; APPLICANT: Reynisdottir, Sigridur Th.
; TITLE OF INVENTION: HUMAN STROKE GENE
; FILE REFERENCE: 2345.2010-003
; CURRENT APPLICATION NUMBER: US/10/067,514
; CURRENT FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: US 09/811/352
; PRIOR FILING DATE: 2001-03-19
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1691139
; TYPE: DNA
; ORGANISM: Human
US-10-067-514-1

Query Match 7.0%; Score 363.6; DB 14; Length 1691139;
Best Local Similarity 80.6%; Pred. No. 7.4e-62;
Matches 464; Conservative 0; Mismatches 104; Indels 8; Gaps 3;

Qy 3387 TGGTGGGTTCTTGGTCTCGCTGACTCCGAAGTAAGAACCGGACGCCCTTGAGGTGAGTGT 3446
Db 213585 TGGTGGGCTCTTGGTCTTCAAGTAAGCGCGGAGCCCTCATGCTGAGTGT 213644

```


Query Match	6.9%	Score 359.8	DB 9	Length 599
Best Local Similarity	81.0%	Pred. No. 7.5e-63		
Matches 490	Conservative 0	Mismatches 87	Indels 28	Gaps 5
QY	3469	TTCAGAGTTTGTTCCTTCAGATGTTCCAGATGTCGCGAGTTTCTCCCTTATCGTCAGTT	3528	
Db	1	TCCGGAGTTTGTTCCTTCAGATGTTCCAGATGTCGCGAGTTTCTCCCTTCATGTGGTT	60	
QY	3529	CGTGGTCTGC--TGACTTCAACAAATGAAGCCGAGACCTTTGCAGTGAGTGTGTGACAGT	3587	
Db	61	CGTAGTCTCACTTGATTTCAGGAGTGAAGCTTCGAGACCTTCGCGAGTGA--GTGTTACAGC	118	
QY	3588	TCTAAAGGCGAGTGGCTGCAGAGTTGTTTCTCCCGGTAGTGTTCGTGGTCTCGCTGA	3647	
Db	119	TCTAAAGGTGGCGGCTCTGGAGTTGTTTCTTCTCCAGTGGTTCGTGGTCTCACTGA	178	
QY	3648	TGTCAGGAATGAAGCTGCAGACCTTCGCGTAAGTGTACAGCTCATAAAGGTAGTGCAA	3707	
Db	179	CTTCAGGGGTGAAGTGCAGACCTTCGC-----AGTGTACAGTTTCATAAGGTAGTGTGC	234	
QY	3708	ACCCAAACAGTAGCAGTAGCAAGATTTATTATGAGAGCAAAAGAACAAAGCTTCCCCA	3767	
Db	235	ACCAAAGGTAGCAGCAGCAAGATTTATTGTGAAGAGCGAAACAAACAAAGCTTCACA	294	

[illegible]

Db 78828 CCACAGTTGTTTCATTCTCCAGTGGGTTTGTGCTCGCTCGCTCAGGAGTGAAGCTG 78887
Qy 3665 CAGACCTCGCGGTAAAGTGTACAGCTCATAAAGGTAGTGCACACCCCAACAGTGAAGCAG 3724
Db 78888 CAGACCTCGGTGAGTGTACAGCTCATAAAGGCAGTGGGACCCCAAGAGTGAACAA 78947
Qy 3725 TAGCAAGATTATTATGAAGAGCAAAAGACAAAGCTTCCCCACCATAGAAACGGACCAG 3784
Db 78948 GAGCAAAATTATTGCAAGAGCAAAAGACAAAGCTCCACACCATGGGAAGAACCCC 79007
Qy 3785 AATTGGTTGCTGCTGCTG---TGGTAGCCAGCTTTTATCCCTTATTTGGC-CACACCCA 3840
Db 79008 TGCAGGTTGCTGCTGCTGCTGGGCGAGCTGCTTTATCCCTTATCTGACTCCACCCA 79067
Qy 3841 CATCTGCTGATTGGCCCAATTACAGAATGCTGATTGGTCCCATTTTATAGCGTGTGAT 3900
Db 79068 CATCTGCTGATTGGCCCAATTATAGAGAGCTGATTGGTCCCATTTTACAGAGAACTGAT 79127
Qy 3901 TGGTGCCTTTT-----TACAGAGTCTGATTGGTGCATTTA 3936
Db 79128 TGGTCCGTTTACAGAGAGCTGATTGGTCTGCTCGACAGGGTGTGATTGGTGCATTTA 79187
Qy 3937 CAATCCCTTTAGCTAGACACAGAGTGTGATTGGTGCCTTTTATAATCCTTTAGCTAGACAC 3996
Db 79188 CAATGCCTGAGCTAGACACAGAGTGTGATTGGTGTATTACAAATCCTCTAGCTAGACGT 79247
Qy 3997 AAAAGTCTACAAAGTCCCCAC 4017
Db 79248 AAAAGTCTCCAAAGTCCCCAC 79268

Search completed: September 23, 2003, 06:41:08
Job time : 1801 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2003, 17:46:24; Search time 9379 Seconds
(without alignments)
13519.187 Million cell updates/sec

Title: US-09-996-956-1

Perfect score: 5217

Sequence: 1 gaacatttaaatatccctc.....taaaaaaaaaaaaaaaaaaa 5217

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_hcc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	579.2	11.1	752	10	BF680192
2	497.4	9.5	501	28	AQ345812
3	434.4	8.3	855	12	BM041081
4	392.6	7.5	821	29	BZ600660

C	5	376.8	7.2	600	28	AQ083508
C	6	376.6	7.2	655	29	AG045844
C	7	371.8	7.1	680	29	AG169322
	8	365.2	7.0	649	2	HS078213
	9	362	6.9	607	28	AQ587990
	10	360.8	6.9	845	28	AQ747612
	11	360	6.9	899	13	BQ926428
	12	358	6.9	630	28	AQ390354
	13	354.4	6.8	22715	28	AQ839851
C	14	354.2	6.8	14771	28	AQ839854
	15	352.8	6.8	617	9	AL704833
	16	352.2	6.8	807	10	BF676760
	17	347	6.7	764	28	AQ742559
	18	342.8	6.6	681	29	AG113408
C	19	341.6	6.5	929	14	CD242806
	20	334.2	6.4	674	28	AQ237458
	21	332.2	6.4	680	28	AQ239957
C	22	331.2	6.3	741	29	AG030015
C	23	330.6	6.3	757	10	BG202513
C	24	325.6	6.2	621	28	B72370
C	25	323.8	6.2	846	14	CD171662
C	26	320.2	6.1	907	13	BQ654152
C	27	320.2	6.1	925	13	BQ645350
C	28	314.6	6.0	636	9	AL603245
C	29	314	6.0	984	13	B0839710
C	30	312.4	6.0	555	28	AQ356187
C	31	307.2	5.9	854	28	AQ897063
C	32	306.2	5.9	711	28	AQ057655
C	33	304.6	5.8	497	28	AQ481794
C	34	302.6	5.8	666	10	BE393102
C	35	299.8	5.7	652	12	BM978961
C	36	297.2	5.7	406	28	B59220
C	37	297	5.7	460	28	AQ539242
C	38	297	5.7	512	28	B59557
C	39	296.2	5.7	556	28	AQ507503
C	40	295.8	5.7	521	28	AQ075172
C	41	294.6	5.6	460	28	B66805
C	42	293.6	5.6	966	13	BX342527
C	43	292.4	5.6	669	28	B91834
C	44	289	5.5	1914	11	AF289565
C	45	287.6	5.5	547	28	AQ053359

ALIGNMENTS

BF680192 602154942F1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4295771 5', linear EST 21-DEC-2000
 mRNA Sequence.
 BF680192 602154942F1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4295771 5', linear EST 21-DEC-2000
 mRNA Sequence.
 BF680192.1 GI:11954087
 EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NIH-MGC http://mgc.nci.nih.gov/.
 1 (bases 1 to 752)
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: CLONETECH Laboratories, Inc.
 cDNA Library Preparation: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
 Plate: LLCML147 row: d column: 12
 High quality sequence stop: 663.

```
FEATURES
source
Location/Qualifiers
1..752
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4295771"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NTH_MGC_83"
/note="Organ: prostate; Vector: pDNR-LIB (Clontech);
Site_1: SfiI (ggcgctcgcc); Site_2: SfiI (ggcattatggcc
); 5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGGCGCCGACATG-dt(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.4 kb (range 0.5-4.0 kb). 14/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."
BASE COUNT 213 a 168 c 174 g 197 t
ORIGIN
Query Match 11.1%; Score 579.2; DB 10; Length 752;
Best Local Similarity 92.8%; Pred. No. 8.8e-11;
Matches 685; Conservative 0; Mismatches 43; Indels 10; Gaps 7;
Qy 173 AATTAGGCGAACCCTTTTGCCCTATGCCGTAACTGTGTCTGCAACTTCTCTTAATTG 232
Db 2 AATTAGGCGAACC-ITTTGCCCTATGCCGTAACTGTGTCTGCAACTTCTCTAA-TG 59
Qy 233 GGAATAGTTAAGCAGATTCATAGAGCTGAATGATATAAATGTACTCGAGATGCATGG 292
Db 60 GGAATAGTTAAGCAGATTCATAGAGCTGAATGATATAAATGTACTCGAGATGCATGG 119
Qy 293 GACTCAAGCTGACCTTATCAAGTCAGATGAGTCTGTCCCTGTCTCAAGCTCGAGCC 352
Db 120 GACTCAAGCTGACCTTATCAAGTCAGATGAGTCTGTCCCTGTCTCAAGCTCGAGCC 179
Qy 353 AATGGTGTATCTTGGCTCACTGCAACCTCCACCTCCAGGTTCAAAAGCTTTCCTGCC 412
Db 180 AATGGTGTATCTTGGCTCACTGCAACCTCCACCTCCAGGTTCAAAAGCTTTCCTGCC 239
Qy 413 TCAGCTCCCAAGTAAGTGGGATTACAGAGGCTTGTGATTTGACACTTCATGATATC 472
Db 240 TCAGCTCCCAAGTAAGTGGGATTACAGAGGCTTGTGATTTGACACTTCATGATATC 299
Qy 473 AGCCAAAGTGAATAAAACAGCTCCTCGAAGAGGACTATGACATCATCAGGTGGGAG 532
Db 300 AGCCAAAGTGAATAAAACAGCTCCTCGAAGAGGACTATGACATCATCAGGTGGGAG 359
Qy 533 TCTCCAGGACAGGGACCTTTTGGAAAGAGCTAGAAAGTGTGAATCTATTAGTCTTC 592
Db 360 TCTCCAGGACAGGGACCC-TTGGAAAGAGCTAGAAAGTGTGAATCTATTAGTCTTC 418
Qy 593 GATATGAATCTCTGCTGTAAAGACTTTCATATTACAGACAGAGGCTATCTCC 652
Db 419 GATATGAATCTCTGCTGTAAAGACTTTCATATTACAGACAGAGGCTATCTCC 478
Qy 653 TAGGCGCAAAAAGTGGCAACAGGCAAGCAGGAGGAA- -GAGATCATGAGCATTC 710
Db 479 TAGGCGCAAAAAGTGGCAACAGGCAAGCAGGAGGAAACACCGAGATCATGAGCATTC 538
Qy 711 AGAGTGCACTGTCCTTTCATATATTCTCAATGCCGTATGTTGGTTTATTTTGGCCAA 770
Db 539 AGAGTGCACTGTCCTTTCATATATTCTCAAGTGCCTGATGTT- -GGTTTCATTGGGCCAA 596
Qy 771 GCATAAATCTGCTCAAGAAAAAATCTCGAGAAAAAAGGTCCTTTCGCAATGT 830
Db 597 GCATGAATCTGCTCAAGAAAAAATCTCGAGAAAAAAGGTCCTTTCGCAATGT 656
Qy 831 TATGTTCTTTTTCACAGCCCTGAGATTTCTGAGGGGAATTCACATATAATGATCAGG 890
Db 657 TATGTT- -CTTTTACCAGGCCCTGAGA- -TTCTGAGGGGAATTCACATATAATGATCAGG 713
```

```
Qy 891 TCATTCAATTACGTTGTG 908
Db 714 CATTCACTTTCGTTGTG 731

RESULT 2
LOCUS AQ345812
DEFINITION RPC111-124L14.TV RPC1-11 Homo sapiens genomic clone RPC1-11-124L14,
genomic survey sequence.
ACCESSION AQ345812
VERSION AQ345812.1 GI:4170708
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 501)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
,J.C.
TITLE Use of BAC End Sequences from Library RPC1-11 for Sequence-Ready
Map Building
JOURNAL Unpublished
COMMENT Other_GSSs: RPC111-124L14.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeetigr.org
Clones are derived from the human BAC library RPC1-11. For BAC
library availability, please contact Pieter de Jong
(pieter@ejong.med.buhalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buhalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: T7
Class: BAC ends.
FEATURES
source
Location/Qualifiers
1..501
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="GB:7547509"
/db_xref="taxon:9606"
/clone="RPC1-11-124L14"
/sex="Male"
/cell_type="Lymphocytes"
/clone_lib="RPC1-11"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC111 Human Male BAC Library"
BASE COUNT 129 a 95 c 121 g 155 t
ORIGIN
Query Match 9.5%; Score 497.4; DB 28; Length 501;
Best Local Similarity 99.6%; Pred. No. 4.4e-08;
Matches 498; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 3254 TTGAGAGAATCTGGCAGAGAAACAATAATTTATCTTACATAAAGCCAGCATTTTA 3313
Db 2 TTGAGAGAATCTGGCAGAGAAACAATAATTTATCTTACATAAAGCCAGCATTTTA 61
Qy 3314 CAGGTTTATTTTAACTATGATTTATCTGATCTTTCATATAAGTGTGCCGG 3373
Db 62 CAGGTTTATTTTAACTATGATTTATCTGATCTTTCATATAAGTGTGCCGG 121
Qy 3374 AATTTATTTCTTGGTGGTTCTTGTCTCGTGTACTTCAAGAAATGAACCCGAGACCC 3433
Db 122 AATTTATTTCTTGGTGGTTCTTGTCTCGTGTACTTCAAGAAATGAACCCGAGACCC 181
Qy 3434 TTGAGGTGAGTGTACAGATTTCTTAAAGATGGTGTTCAGATTTGTTCTTCAGATGTT 3493
Db 182 TTGAGGTGAGTGTACAGATTTCTTAAAGATGGTGTTCAGATTTGTTCTTCAGATGTT 241
```

QY	3494	CAGATGTGCGGAGTTTCTCCCTTATGGTGAGTTGCGTCTCGCTGACTTCAACAATG	3553
Db	242	CAGATGTGCGGAGTTTCTCCCTTATGGTGAGTTGCGTCTCGCTGACTTCAACAATG	301
QY	3554	AAGCCGACACCTTTCGAGTGAGTGTGCACAGTCTTTAAAGGCGAGTGCCTCGAGAGTTG	3613
Db	302	AAGCCGACACCTTTCGAGTGAGTGTGCACAGTCTTTAAAGGCGAGTGCCTCGAGAGTTG	361
QY	3614	TTTGTTCCTCCCGTAGTTCCTGGTCTCGTGTGATGTGAGTGTGAGTGTGAGTGTGAGTGTG	3673
Db	362	TTTGTTCCTCCCGTAGTTCCTGGTCTCGTGTGATGTGAGTGTGAGTGTGAGTGTGAGTGTG	421
QY	3674	CGCGTAAAGTGTACAGTCTCATAAAGTGTGATGTGAGTGTGAGTGTGAGTGTGAGTGTG	3733
Db	422	CGCGTAAAGTGTACAGTCTCATAAAGTGTGATGTGAGTGTGAGTGTGAGTGTGAGTGTG	481
QY	3734	TTATTATGAAGAGCAAGA	3753
Db	482	TTATTATGAAGAGCAAGA	501
RESULT 3			
BM041081/c			
LOCUS			
DEFINITION	60361481F1 NTH_MGC_108 Homo sapiens cDNA clone IMAGE:5556482 5', mRNA sequence.	855 bp	linear EST 07-NOV-2001
ACCESSION	BM041081		
VERSION	BM041081.1	GI:16770336	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 855)		
TITLE	NIH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: ATCC		
FEATURES	cdna Library Preparation: Ling Hong/Rubin Laboratory cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: L1CM2020 row: f column: 03 High quality sequence stop: 819. Location/Qualifiers		
SOURCE	1. .855 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:5556482" /tissue_type="Wilms' tumor, cell line" /lab_host="DH10B (phage-resistant)" /clone_lib="NIH_MGC_108" /note="Organ: kidney; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."		
BASE COUNT	244 a	214 c	203 g
ORIGIN			194 t
Query Match	8.3%	Score 434.4;	DB 12; Length 855;
Best Local Similarity	81.6%;	Pred. No. 2.8e-06;	
Matches 563; Conservative	0;	Mismatches 96;	Indels 31; Gaps 4;


```
Matches 524; Conservative 0; Mismatches 98; Indels 54; Gaps 4;
Qy 3406 CTGACTCAAGAATGAACCGCAGACACCTTGAGTGTAGTGC-ACAGTCTCTTAAGATGG 3464
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
680 CTGACTTCAGAATGAAGCGCAGACCCCTCGTGTGAGTGTACAGTCTCTTAAGATGG 621
Qy 3465 TGTGTTCCAGATTTGTTCTTCAGATGTTTCAGATGTCTCGGAGTCTTCCTCCATGATG 3524
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
620 TGTGTCCAGAGTTTGTCTCTCTGATGTTTCAGATATGCTGAGTCTTCTCTCTCTG 561
Qy 3525 AGTTCGTTGCTCGTCTGACTTCACATGAGTGAAGCGCAGACCTTTCAGTGTGAGTGC 3584
Db || || || || || || || || || || || || || || || || || || || || ||
560 GGTTCGGGCTTCTGACTTCAGTGAAGTGAAGTGCAGACCTTCACAGTGA--GTGTAC 503
Qy 3585 AGTTCCTTAAAGGAGTGCCTCGCAGAGTGTGTTGCTCCCGGTAGTGTGCTGCTCGC 3644
Db || || || || || || || || || || || || || || || || || || || || ||
502 AGCTCTTAAAGTGTGCTGCTGAGTGTGTTCAATCTCTCTGTTGGTGTCTGTTCTC 443
Qy 3645 TGATGTCAGGAATGAAGTGCAGACCTCGCGTAACTGTATACAGCTCATAAAGGTAGT 3704
Db || || || || || || || || || || || || || || || || || || || || ||
442 TGACTTCAGGAGTGAAGCCACAGACCTTCGAGAAAGTGTACAGCTCATAAAGGTAGT 383
Qy 3705 CAAACCAACAGTGAAGTGCAGATGATTCAGATTTATTAAGAGCAAAAGCAAGCTTC 3764
Db || || || || || || || || || || || || || || || || || || || || ||
382 CAGACCCAAAGAGTGAAGCAGATGATTTATGTGAAGAGTGAAGAGCAACAGCTTC 323
Qy 3765 CCACCATAGAAAGCAGCAAGTGTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3821
Db || || || || || || || || || || || || || || || || || || || || ||
322 ACAGCATGGAAGGTAGTGTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 263
Qy 3822 CCTATTGTCACACACACACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3881
Db || || || || || || || || || || || || || || || || || || || || ||
262 CCTATTGTCACACACACACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 225
Qy 3882 CATTTATAGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3941
Db || || || || || || || || || || || || || || || || || || || || ||
224 -----TTTATACAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 191
Qy 3942 CTTTAGTACACACAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4001
Db || || || || || || || || || || || || || || || || || || || || ||
190 CTTTAGTACACACAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 131
Qy 4002 TTTACAGTGTCCACACACACACACACACACACACACACACACACACACACACACAC 4061
Db || || || || || || || || || || || || || || || || || || || || ||
130 TTTCTCAAGTGTCCACACACACACACACACACACACACACACACACACACACACACAC 71
Qy 4062 AGGTTCAACCAAGTTT 4077
Db || || || || || || || || || || || || || || || || || || || || ||
70 AACCTTAACTTACTTT 55

RESULT 8
HSM078213
ID HSM078213 standard; RNA; EST; 649 BP.
XX AC BX507549;
XX SV BX507549.1
XX
DT 09-MAY-2003 (Rel. 75, Created)
DT 09-MAY-2003 (Rel. 75, Last updated, Version 1)
XX
DE Homo sapiens mRNA; EST DKFZp686i1625_r1 (from clone DKFZp686i1625)
XX
XX EST; expressed sequence tag.
XX
OS Homo sapiens (human)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX
RN [1]
RP Pouskta A., Albert R., Moosmayer P., Schupp I., Wellenreuther R.,
RA 1-649
```

[illegible]


```

Db      702 GCTGATTGTCCTCATTTTACAGAGCGCTGAGTGGTGGCTTTTATACACAGTGGCTGATTGCT 761
QY      3930 GCATTATCAATCCTTTAGCTACACAGAGTGTCT-GATTGGTGGCTTTATATATCCTTTAG 3988
Db      762 GCATTANCATCCTTTAGCTAGACCAGAGTGTGATTTGGTCATTTTATACACTTTTAG 821
QY      3989 CTAGACA 3995
Db      822 ACACAGA 828

RESULT 12
AQ390354
LOCUS   CITBI-E1-2545M11.TR CITBI-E1 Homo sapiens genomic clone 2545M11,
DEFINITION genomic survey sequence.
ACCESSION AQ390354
VERSION   AQ390354
KEYWORDS  AQ390354.1 GI:4361377
SOURCE    GSS.
ORGANISM  Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1. (bases 1 to 630)
AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
          Venter,J.C.
TITLE     Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
          Map Building
JOURNAL   Unpublished
COMMENT   Other_GSSs: CITBI-E1-2545M11.TF
          Contact: Shaying Zhao, William Nierman, Mark Adams
          Department of Eukaryotic Genomics
          The Institute for Genomic Research
          9712 Medical Center Dr., Rockville, MD 20850
          Tel: 301 838 0200
          Fax: 301 838 0208
          Email: hbe@tigr.org
          Clones are available from Research Genetics (info@resgen.com). BAC
          end search page:
          http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html.
          Seq primer: M13 Reverse
          Class: BAC ends.
FEATURES             Location/Qualifiers
     source           1..630
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="2545M11"
                     /sex="male"
                     /cell_type="sperm"
                     /clone_lib="CITBI-E1"
     note="Vector: pBelobAC11; Site_1: EcoRI; Site_2: EcoRI;
          Caltech Human BAC Library D"
BASE COUNT  144 a 134 c 175 g 177 t
ORIGIN

Query Match          6.9%; Score 358; DB 28; Length 630;
Best Local Similarity 83.0%; Pred. No. 0.00088;
Matches 448; Conservative 0; Mismatches 80; Indels 12; Gaps 3;

QY      3356 TCATATAAGTGTCCCGGAATTTATTTCTTGTGGGTCTTGTGCTCGCTGACTCCAA 3415
Db      98  TCACCAAAATGTGTCGGGAATTTATCTTCCAGTGGGTCTTGTGCTCGCTGACTCAA 157
QY      3416 GAATGAACCCGACACCCCTTGAGTGAGTGCATCAGTTCTTAAAGATGGTGTTCAGAG 3475
Db      158 GAATGAAGCCATGGACCCCTTGAGTGAGTGTACAGCTCTTAAAGATGGTGTGTCGGAG 217
QY      3476 TTTGTTCTTCAGATGTTTCAGATGTGTCGGAGTTTCTCCCTTATGTTAGTTCGTGTC 3535
Db      218 TTTGTTCTTCAGAGATGTTTCAGATGTATCCAGAGTTTCTTCCTTCTGGTGGTTCGTGTC 277
QY      3536 TCGCTGACTTCAACAATGAAGCCGACACCTTTTGCAGTGAGTGTGTGACAGTTCCTTAAAG 3595

```

```

Db      278 TTGTGTGACTTCAGGAGTGA-----AGACCTTTGCAGTGA--GTGTTACAGCTCTTAAAG 329
QY      3596 CGAGTGGCTGCAGAGTGTGTTGTTCTCCCGTAGTGTTCGGTGTCTCGCTGATGTCAGGA 3655
Db      330 GTGGTGGCTGCAGAGTGTGTTGTTCTCCCGTAGTGTTCGGTGTCTCGCTGATGTCAGGA 389
QY      3656 ATGAAGCTGCAGACCCCTCGCGGTAAAGTGTTCACAGCTCATAAAGTAGTGCAAACCCCAAC 3715
Db      390 ATGAAGCCACAGACTCTTGCAGTGAGTGTTCAGAGCTCATAAAGTAGTGCAGCCCAAG 449
QY      3716 AGTGAGCAGTAGCAAGATTTATTTGAAGAGCAAAAGAAACAAAGCTTCCCCACCATAGAA 3775
Db      450 AATGAGCAGCACCAAGATTTATTTGAAGAGCAAAAGAAACAAAGCTTCAAAGCGTGGAA 509
QY      3776 ACGGACCAAGATTTGTTGCTGCT---GCTGTGGTAGCCAGCTTTTATTCCTCTTATTTGG 3831
Db      510 GGGGACCCGAGCGGTGCTGCTGCTGCTGCTGAGTGGCCAGCTTTTATTCCTCTTATTTGT 569
QY      3832 CCACACCCACATCCTGCTGCTGATTTGGCCCATTTTACAGAAATGCTGATTGCTCCATTTATAG 3891
Db      570 CCCCGTCCACATCCTGCTGATTTGTCATTTTACAGAGTGTGATTGGTCCATTTTCATAG 629

RESULT 13
AQ839851
LOCUS   69115-C79 CITB Homo sapiens genomic clone 69115, genomic survey
DEFINITION sequence.
ACCESSION AQ839851
VERSION   AQ839851.1 GI:6652483
KEYWORDS  GSS.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1. (bases 1 to 22715)
AUTHORS   Carpten,J.D., Bonner,T.I., Smith,J.R., Faruque,M.U., Stephan,D.A.,
          Connors,T.D., Makalowska,I., Robbins,C.M., Scott,N., Sood,R.,
          Pinkett,H., Morgenbesser,S.D., Su,K., Graham,C., Gregory,S.G.,
          Williams,H., McDonald,L., Baxevasanis,A.D., Klingler,K.W. and Landes
          ,G.M.
TITLE     A 6-Mb high-resolution physical and transcription map encompassing
          the hereditary prostate cancer 1 (HPC1) region
JOURNAL   Genomics 64 (1), 1-14 (2000)
MEDLINE   20175426
PUBMED    10708513
COMMENT   Contact: Carpten JD
          Cancer Genetics Branch
          National Human Genome Research Institute/National Institutes of
          Health
          Bldg. 36, Room 3D04, 36 Convent Drive, Bethesda, MD
          Tel: 301 435 5626
          Fax: 301 435 5465
          Email: jdc@nhgri.nih.gov
          Class: Shotgun.
FEATURES             Location/Qualifiers
     source           1..22715
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="69115"
                     /clone_lib="CITB"
BASE COUNT  7023 a 4136 c 4231 g 7322 t 3 others
ORIGIN

Query Match          6.8%; Score 354.4; DB 28; Length 22715;
Best Local Similarity 82.2%; Pred. No. 7.9e-05;
Matches 457; Conservative 0; Mismatches 91; Indels 8; Gaps 4;

QY      3504 CGGAGTTTCTCCCTTATGTTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3563
Db      20272 CTGGATTGGTCTTCCAGTGGTGTCTTGGTCTCACTGACTTCAAGAGTGAAGCCGCTGA 20331

```

```
QY 3564 CTTTTCAGTGAAGTGTGTGACAGTCTTAAAGGAGTGCCTGACAGAGTGTGTTGTCCTC 3623
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20332 CCCTCGGGTGA--GTGTACAGTCTTAAAGATGGCGCGTCTGGAGTTTTTTGTTCTC 20389
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3624 CCGTAGTGTCTGCTCGC-TGATGTCAGGAATGAAGCTGCA-GACCCCTCGCGGTAAG 3681
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20390 CCAGTGGTTGTGCTCGCTGACCTCAGAGTGAAGCGGAGGACCTTCGCACATCA 20449
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3582 TGTACAGCTCATAAAGTAGTGCACAAACCAACAGTGCAGTAGTGAAGATTTATATG 3741
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20450 TGTACAGCTCATAAAGTAGTGCAGACCCCAACAGTGCAGCAGCAAGATTTACGGTG 20509
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3742 AAGAGCAAAAGCAAAAGCTTCCCCACCATGAACAGCAGCAAGATTTGTTGCTGCT 3801
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20510 AAGAGCAAAAGCAATAGCTTCCACAGCAGGAAGGAGCCGCGCATGCTGCCACATGCT 20569
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3802 G----TGGTAGCCAGCTTTTATTCCTTATTTGCCCCACACCCACATCCTGCTGATGGCC 3857
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20570 GGCTCGGTAGCCAGCTTTTATTCCTTATTTGGCCCTGCCCATGTACTGCTGATGGTC 20629
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3858 CATTTTACAGATGCTGATGGTCCATTTATAGCGTGTGATGGTGGTGGCTTTTACAGA 3917
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20630 CATTTTACAGAGCTGATGGTCCATTTTACACAGTGTGATGGTGGTGGTGGTGGTGGT 20689
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3918 GTGCTGATGCTGATTTACAATCCTTTAGCTAGACACAGAGTGTGATGGTGGTGGCTTTA 3977
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20690 GTGCTGATGCTGATTTACAACCTTTAGCTAGACACAGAGTGTGATGGTGGTGGTGGT 20749
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3978 TAATCCTTTAGCTAGACACAAAGTTTACAAGTCCCCACCCACCCAGAGTCCCGCTG 4037
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20750 CAATCCTTTAGCTAGACACAAAGTTTCCCAAGTCCCCACCCACCCAGAGTCCCGCTG 4089
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4038 GCTTCACCTCTCGTAA 4053
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20810 GCTTCACCTCTCTCTTA 20825
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
LOCUS A0839854/C 14771 bp DNA linear GSS 31-MAR-2000
DEFINITION 260L13-C54 CITB Homo sapiens genomic clone 260L13, genomic survey
sequence.
ACCESSION A0839854
VERSION A0839854.1 GI:6652486
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
  1 (bases 1 to 14771)
  Carpen, J.D., Makalowska, I., Robbins, C.M., Scott, N., Sood, R.,
  Connors, T.D., Bonner, T.I., Smith, J.R., Faruque, M.U., Stephan, D.A.,
  Pinkett, H., Morgenbesser, S.D., Su, K., Graham, C., Gregory, S.G.,
  Williams, H., McDonald, L., Baxevanis, A.D., Klingler, K.W. and Landes
  , G.M.
  A 6-Mb high-resolution physical and transcription map encompassing
  the hereditary prostate cancer 1 (HPC1) region
  Genomics 64 (1), 1-14 (2000)
  20175426
  10708513
COMMENT Contact: Carpen JD
  Cancer Genetics Branch
  National Human Genome Research Institute/National Institutes of
  Health
  Bldg. 36, Room 3D04, 36 Convent Drive, Bethesda, MD
  Tel: 301 435 5626
  Fax: 301 435 5465
  Email: jdc@nhgri.nih.gov
  Class: shotgun.
  Location/Qualifiers
  1..14771
  /organism="Homo sapiens"

/mot_type="genomic DNA"
/db_xref="taxon:9606"
/clone_lib="CITB"
BASE COUNT 4135 a 2762 c 2928 g 4937 t 9 others
ORIGIN
Query Match 6.8%; Score 354.2; DB 28; Length 14771;
Best Local Similarity 77.08; Pred. No. 0.00011;
Matches 532; Conservative 0; Mismatches 103; Indels 56; Gaps 6;
QY 3358 ATATAAGTGTCCCGGAAATTTATTTCTTCTGGTGGTCTTCTGGTCTCGCTGACCTCCAGA 3417
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2239 ATACATTTGTCTCCAGAGTTTATGCCCTTCCAGTGGTTCTTCTGCTCGCTGACCTCAAGA 2180
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3418 ATGAACCGGAGACCCCTTGAGTGTGACAGTCTTCTTAAAGATGGTGTGTTTTCAGAGTT 3477
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2179 ACGAAGCGGAGAACTTCACAGTGTGAGTCTTAAAGATGGTGTGTCGCGGAAT 2120
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3478 TGTTCCTTCAGATGTTCCAGATGTGTCGGAGTGTTCCTTATGTTGAGTTCGTGGTCTC 3537
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2119 TGTTCCTTCAGAGTGTCTA-----GGAGTTTCTTCCCTTCCGCTGAGTTAGTGGTCTC 2067
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3538 GCTGACTTCAACAAATGAAGCCGACACCTTTTGCAGTGTGTGACAGTTCCTTAAAGGC 3597
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2066 GCTGACTTCAGGAGTGAAGCTGTGGACCTTCGCAGTGAGT--GTTACAGCTCTTTAAAGGT 2009
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3598 AGTCCGTCACAGATGTTGTTCTCTCCGGTAGTGTGCTGCTCGCTGATGTGAGGAAT 3657
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2008 GCGCGCTCTGGAGTGTGTTGGCCCTCTCTGTGGGTTCATGGTCTGGGTGATCTCAGGAAT 1949
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3658 GAAGCTCAGACCCCTCGGCTGAAGTGTACAGCTCATAAAG---GTAGTGCACAAACCCAAA 3714
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1948 GAAGCTCAGACCCCTGATGTTGAGTGTACAGCTCATAAAGTGTAGTGCACCAACCA 1889
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3715 CAGTGCAGCTAGCAAGATTTATTTATGAAGAGCAAAAGAAAGCTTCCGCCACATAGA 3774
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1888 GAGTGAGCAGCAGCAAGATTTATTTGAGGAGCAAAACAAAGCTTCCACAGCATAGA 1829
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3775 AACGGACCAAGTTGTTGCTGCTG---CTGTGTAGCCAGCTTTTATCCCTTATTTG 3830
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1828 AGTGGAGCAGAAAGAGCTGGCCCTCTGGGTGGCAGCTTTTATCCCTTATTTG 1769
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3831 GCCACACCCACATCTCTGATTTGGCCCATTTTACAGAAAGCTGATTTGGTCCATTTAT- 3889
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1768 GCCCTGCCACATCTTACTGTTGTTCCATTTTACAGAGTCTGATTTGGTCCATTTTTC 1709
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3890 AGCGTGTGATTTGTTGCTGCT-----T 3910
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1708 ACAGTGTGACTGTGTCATTTTACAATCCTTTAGCTAGACACAGAGCAGTGTGTTGTCAT 1649
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3911 TTACAGAGTCTGATTTGGTGCATTTACAATCCTTTAGCTAGACACAGAGTGTGCTGATTTG 3970
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1648 TTTTAGAGTCTGACTGGTGCATTTTACAATCCTTTAGCTAGACACAGAGGCTGATTTGGT 1589
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3971 GCCTTTATAATCCTTTAGCTAGACACAAAG 4001
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1588 AAATTTACATCCTTTAGCTAGACAAAGAG 1558
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
LOCUS AL704833
DEFINITION DKFZp686B2033_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
ACCESSION DKFZp686B2033 5', mRNA sequence.
VERSION AL704833
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
  1 (bases 1 to 617)
```

AUTHORS Ottenwaelder, B., Obermaier, B., Mewes, W., Mewes, H.W., Weil, B. and Wiemann, S.
TITLE EST (Ottenwaelder, B., Obermaier, B., Mewes, H.W., Weil, B. and Wiemann, S.)
JOURNAL Unpublished
COMMENT Contact: Ottenwaelder B
MIPS Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by Medigenomix (Martinsried/Germany) within the cDNA sequencing consortium of the German Genome Project. No sl sequence available.
This clone (DKFZp686B2033) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin- Charlottenburg, GERMANY; Email: cloneerzpd.de.
Location/Qualifiers
1..617

FEATURES
source
136 a 147 c 163 g 171 t
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686B2033"
/tissue_type="human skeletal muscle"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="686 (synonym: hlcc3)"
/note="Vector: pTriplex2; Site_1: SfiIA; Site_2: SfiIB;
cDNA collection"
BASE COUNT 136 a 147 c 163 g 171 t
ORIGIN

Query Match 6.8%; Score 352.8; DB 9; Length 617;
Best Local Similarity 80.3%; Pred. No. 0.0013;
Matches 490; Conservative 0; Mismatches 82; Indels 38; Gaps 5;
QY 3448 ACAGTTCCTAAAGATGGTGTTCAGAGTTGTTCCCTCAGATGTCAGATGTGTCGGGA 3507
DB 2 ACAACTCTAAAGATGGTGTTCAGAGTTGTTCCCTCAGATGTCAGATGTGTCGGGA 61
QY 3508 GTTTCCTCCCTGATGTTGGTGTCTCGCTGACTTCACAATGAAGCCGACACCTT 3567
DB 62 ---ACTTCTTCGGTGGGTTCGTGGTCTCTACTGACTTCAGAGTGAAGCCGACATCTT 118
QY 3568 TGCAGTAGTGTGCACAGTTCTTAAAGGCAGTGGCTCAGAGTTGTTTGTTCCTCCGG 3627
DB 119 CGCACTGA -GTGTTACAGCTCTTAAAGGTGSCAGTCAGGAGTTGTTGTTCTCCGG 176
QY 3628 TAGGTTCTGGTCTCGCTGATGTCAGGAATGAAGTCGACCCCTCGCGGTAAGTGTAC 3687
DB 177 TGGGTTTGGTCTCGCTGACTTCAGGAATGAAGCCGCGGCTTGTGGTGTGTTAC 236
QY 3688 AGCTCATAAAGTGTGCAACCAACAGTGAAGAGTGAAGATTTATTATGAAGAGC 3747
DB 237 AGCTCATAAAGTGTGACAGCCCGAAGAGTGAAGAGTGAAGATCTACTGTGAAGAGC 296
QY 3748 AAAGAACAAGCTTCCCAACCATAGAACCGACAGAAATGGTTGCTGCTGCTGT --- 3803
DB 297 AAAGAACAAGCTTTCACAGCGTGAAGGGAGCCAGCGGGTGTGCTGCTGCTGCTG 356
QY 3804 GGTACCCAGCTTTTATTCCTTATTTGGCCACACCCATCCTGCTGATTTGCCCATTTT 3863
DB 357 GGTGCCAGCTTTTATTCCTTATTTGGCCCTGCCACATCCTGCTGTTGGTCCATCTT 416
QY 3864 ACAGATGCTGATGCTCATTTTATAGCTGCTGATTTGGTGGTGGTGGTGGTGGTGGT 3923
DB 417 ACAGAGCACTGACTGGTCCA -----TTTACAGAGTGGTG 451
QY 3924 ATTGGTGATTTACAAATCTTTAGTACACAGAGTGTGATTTGGTGGTGGTGGTGGTGGT 3983
DB 452 ATTGGTGCTTACAAATCCTT ---TAAACAGAGAGCTGATTTGGTGCATTTACAATCC 507
QY 3984 TTTAGCTAGACAAAGAAGTCTTACAAGTCCCAACCCAGAGCTCCGCTGGCTTCA 4043

Db 508 TTTAGCTAGACAAAGAAGTCTTCCAAGTCCCAACCCAGAGTCCAGCTGGCTTCA 567
QY 4044 CCTCTCGTAA 4053
DB 568 CCTCTCATTA 577

Search completed: September 23, 2003, 01:55:31
Job time : 9389 secs

THIS PAGE BLANK (0.5PT)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 22, 2003, 17:01:44 ; Search time 18295 Seconds
(without alignments)
11665.784 Million cell updates/sec

Title: US-09-996-956-1

Perfect score: 5217

Sequence: 1 gaaacttaaatatccctc.....taaaaaaaaaaaaaaaaaaaaaa 5217

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_em:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htg_mus:*
- 34: em_htg_pln:*
- 35: em_htg_rod:*
- 36: em_htg_mam:*
- 37: em_htg_vrt:*
- 38: em_sy:*
- 39: em_htgo_hum:*
- 40: em_htgo_mus:*
- 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Length	DB	ID	Description
1	5217	100.0	5217	6	AX467721	Sequence
2	4957	95.0	5092	6	AX467723	Sequence
3	4763.6	91.3	168595	9	AC013401	AX467723 Homo sapi
4	4758.4	91.2	32463	6	AX467725	Sequence
5	1458.4	28.0	1603	9	AF223389	Homo sapi
6	514.8	9.9	131458	9	AC002386	Homo sapi
7	510.4	9.8	193041	9	AC092017	Homo sapi
8	503	9.6	162126	9	AL354711	Human DNA
9	503	9.6	297235	2	AL499603	Homo sapi
10	494.6	9.5	99632	2	AL157779	Human DNA
11	484.4	9.3	123637	9	HS3067	Human DNA
12	484.2	9.3	99527	9	AL137015	Human DNA
13	484.2	9.3	135090	9	HS287614	Human DNA
14	483.8	9.3	159797	9	AC104697	Homo sapi
15	482.4	9.2	179598	9	CNS01DRO	Human chr
16	482	9.2	46975	9	AL359376	Human DNA
17	482	9.2	68697	9	AL451009	Human DNA
18	480.4	9.2	145993	9	AC098972	Homo sapi
19	480.4	9.2	158508	9	AC107622	Homo sapi
20	479.4	9.2	159851	9	AC036114	Homo sapi
21	479.4	9.2	164009	9	AC079064	Homo sapi
22	479.4	9.2	195344	2	AC025248	Homo sapi
23	479.2	9.2	76028	9	AL451129	Human DNA
24	477.8	9.2	117583	9	AC034238	Homo sapi
25	477.8	9.2	133322	2	AC091845	Homo sapi
26	476.4	9.1	203076	9	AL591074	Human DNA
27	476.2	9.1	183494	9	AC091977	Homo sapi
28	470.8	9.0	103728	9	AL445219	Human DNA
29	470.4	9.0	3265	9	HSAC001178	Homo sapi
30	470.4	9.0	62466	9	AC005613	Homo sapi
31	468.4	9.0	188818	2	AC098674	Homo sapi
32	467.4	9.0	144626	9	AL159159	Human DNA
33	467	9.0	114016	9	AL365396	Human DNA
34	466.8	8.9	52813	9	AC010991	Homo sapi
35	465.8	8.9	32213	2	AC012120	Homo sapi
36	465.8	8.9	126253	9	AP002023	Homo sapi
37	465.8	8.9	127143	2	AL138736	Homo sapi
38	465.8	8.9	173933	9	AP002022	Homo sapi
39	465.2	8.9	222883	2	AC092847	Homo sapi
40	465	8.9	142234	9	AC110620	Homo sapi
41	465	8.9	171196	2	AP001369	Homo sapi
42	463.2	8.9	157604	9	AC021093	Homo sapi
43	463.2	8.9	157633	9	AC123776	Homo sapi
44	463	8.9	108182	9	AC009308	Homo sapi
45	462	8.9	133614	2	AC025433	Homo sapi

ALIGNMENTS

RESULT 1
AX467721
LOCUS AX467721 5217 bp DNA linear PAT 16-JUL-2002
DEFINITION Sequence 1 from Patent WO0244332.
ACCESSION AX467721
VERSION AX467721.1 GI:21900893
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1
AUTHORS Sun, Z., Li, X. and Jay, G.
TITLE Prostate polynucleotides and uses
JOURNAL Patent; WO-0244332-A-1-06-JUN-2002;

[illegible]

Db	3061	TCCATGAACCTGGTAAACTGGTACTAGTATCTCTATTGGACATCTGGAGAAACCAATGGGA	3120
Qy	3121	GTTGATAAACAGTAGAGTTAAAAATTACTCTTCATATATTATATATGCGCTCAATCTCACAG	3180
Db	3121		
Db	3121	GTTGATAAACAGTAGAGTTAAAAATTACTCTTCATATATTATATGCGCTCAATCTCACAG	3180
Qy	3181	ACATCTCTGCTACAAAAGCTATCAATCTATAGATATGCGGCATATAGATGACCTTGGGGC	3240
Db	3181		
Db	3181	ACATCTCTGCTACAAAAGCTATCAATCTATAGATATGCGGCATATAGATGACCTTGGGGC	3240
Qy	3241	ACACTAGAAATCTTTTGAGAGAAATCTGGCAGAGAAAACAAATATTTATTCCTACAATAAA	3300
Db	3241		
Db	3241	ACACTAGAAATCTTTTGAGAGAAATCTGGCAGAGAAAACAAATATTTATTCCTACAATAAA	3300
Qy	3301	ACCCAGCATTTTACAGCTTTTATTTTTAACTATGAAGTATTTGTTATCTGTATCTTTCATA	3360
Db	3301		
Db	3301	ACCCAGCATTTTACAGCTTTTATTTTTAACTATGAAGTATTTGTTATCTGTATCTTTCATA	3360
Qy	3361	TAAGTGTGCCGGAAATTTATTTCTCTGTGTGGGTTCTTGTGTCTCGCTGACTCCAAGATG	3420
Db	3361		
Db	3361	TAAGTGTGCCGGAAATTTATTTCTCTGTGTGGGTTCTTGTGTCTCGCTGACTCCAAGATG	3420
Qy	3421	AAACCCGAGACCTTGTAGGTGAGTGTACAGTTCTTTAAAGATGGTGTGTTCACAGTTTGT	3480
Db	3421		
Db	3421	AAACCCGAGACCTTGTAGGTGAGTGTACAGTTCTTTAAAGATGGTGTGTTCACAGTTTGT	3480
Qy	3481	TCCTTCAGATGTTTCAGATGTGTCGGAGTTTCTCCCTCTATGTTGAGTTCGTGTGCTCGCT	3540
Db	3481		
Db	3481	TCCTTCAGATGTTTCAGATGTGTCGGAGTTTCTCCCTCTATGTTGAGTTCGTGTGCTCGCT	3540
Qy	3541	GACTTCAACAATGAAGCCGAGACCTTTGTCAGTGTAGTGTGTGACAGTTCTTAAAGCGAGT	3600
Db	3541		
Db	3541	GACTTCAACAATGAAGCCGAGACCTTTGTCAGTGTAGTGTGTGACAGTTCTTAAAGCGAGT	3600
Qy	3601	GGGTCCAGAGTCTTTGTTCTCTCCCGTAGGTTTCGTGCTCGCTGATGTCAGGAATGAA	3660
Db	3601		
Db	3601	GGGTCCAGAGTCTTTGTTCTCTCCCGTAGGTTTCGTGCTCGCTGATGTCAGGAATGAA	3660
Qy	3661	GCTGCACACCTCGCGTAAGTCTTACAGCTCATAAAGTAGTGTGCAACCCCAACACAGTGA	3720
Db	3661		
Db	3661	GCTGCACACCTCGCGTAAGTCTTACAGCTCATAAAGTAGTGTGCAACCCCAACACAGTGA	3720
Qy	3721	GCAGTAGCAAGATTTATATGAGAGACAAAGAACAAAGCTTCCCAACCATATGAAACGGGA	3780
Db	3721		
Db	3721	GCAGTAGCAAGATTTATATGAGAGACAAAGAACAAAGCTTCCCAACCATATGAAACGGGA	3780
Qy	3781	CCAGAAATGTGTGCTGCTGTGTGTAGCCAGCTTTTATTCCTCTTATTTGGCCACACCCCA	3840
Db	3781		
Db	3781	CCAGAAATGTGTGCTGCTGTGTGTAGCCAGCTTTTATTCCTCTTATTTGGCCACACCCCA	3840
Qy	3841	CATCCTGCTGATTGGCCATTTTACAGATGCTGATTGGTCCCATTTTATAGCGTGTCTGAT	3900
Db	3841		
Db	3841	CATCCTGCTGATTGGCCATTTTACAGATGCTGATTGGTCCCATTTTATAGCGTGTCTGAT	3900
Qy	3901	TGTTGCGTTTTTACAGAGTCTGATTGGTGCATTTTACAATCTTTTACAGTGTAGACACAGAGT	3960
Db	3901		
Db	3901	TGTTGCGTTTTTACAGAGTCTGATTGGTGCATTTTACAATCTTTTACAGTGTAGACACAGAGT	3960
Qy	3961	GTGTGATTGGTGCCTTTATATTCCTTTAGCTAGACACAAAGATTCTACAGTCCCAACCCCA	4020
Db	3961		
Db	3961	GTGTGATTGGTGCCTTTATATTCCTTTAGCTAGACACAAAGATTCTACAGTCCCAACCCCA	4020
Qy	4021	ACCAGAAGCTCGCTGGCTTACCTCTCCGTAGGAAATTTAGAGTTTCAACAAAGTTTCAA	4080
Db	4021		
Db	4021	ACCAGAAGCTCGCTGGCTTACCTCTCCGTAGGAAATTTAGAGTTTCAACAAAGTTTCAA	4080
Qy	4081	AGTGCTAAAACTACAGTTTCTCATCTCTCGCAACTGGATTTCCTCACTCATGTGTTGAATC	4140
Db	4081		
Db	4081	AGTGCTAAAACTACAGTTTCTCATCTCTCGCAACTGGATTTCCTCACTCATGTGTTGAATC	4140
Qy	4141	CCAGGCTCTAAGACTTAACTTGCAATCTCTGTGACTTTTATGTTCTCGCAATTTACACAAG	4200
Db	4141		
Db	4141	CCAGGCTCTAAGACTTAACTTGCAATCTCTGTGACTTTTATGTTCTCGCAATTTACACAAG	4200

QY		4201	CTACTATCTGTCA	CAATCTCTGGTGT	TAACTTCAGAC	TAAACTTCTTTT	GATTCAACA	ATG	4260
Db		4201	CTACTATCTGTCA	CAATCTCTGGTGT	TAACTTCAGAC	TAAACTTCTTTT	GATTCAACA	ATG	4260
QY		4261	ACCACACACTTTT	TGGTGTAGGTTT	TGCTATCGGTTT	TATGTACTGTTT	TAATAGAGCT		4320
Db		4261	ACCACACACTTTT	TGGTGTAGGTTT	TGCTATCGGTTT	TATGTACTGTTT	TAATAGAGCT		4320
QY		4321	TCCTCCAGAAAT	TGGTAGATGA	AGAGGAAGTAG	CACATTCCTA	AAAAATGTA	ACCATGC	4380
Db		4321	TCCTCCAGAAAT	TGGTAGATGA	AGAGGAAGTAG	CACATTCCTA	AAAAATGTA	ACCATGC	4380
QY		4381	CTTTTCAAGTCA	CAAGCATCCCT	ATACATGGCTG	CTCAAGGGTGG	CTCAGAAATAG	GTAGAG	4440
Db		4381	CTTTTCAAGTCA	CAAGCATCCCT	ATACATGGCTG	CTCAAGGGTGG	CTCAGAAATAG	GTAGAG	4440
QY		4441	TTAGAATTTAA	AGTAATTTGGT	GTAGCGATGA	AGAGCTTCA	CTAAAAAGCT	TATATAC	4500
Db		4441	TTAGAATTTAA	AGTAATTTGGT	GTAGCGATGA	AGAGCTTCA	CTAAAAAGCT	TATATAC	4500
QY		4501	ATCAACTGAA	ATGTAATTA	ATTTGGAACAT	TTTCCAGGCAT	CCCTGTTAT	TTTCTGCT	4560
Db		4501	ATCAACTGAA	ATGTAATTA	ATTTGGAACAT	TTTCCAGGCAT	CCCTGTTAT	TTTCTGCT	4560
QY		4561	CTCTTTCTCT	TGCTTGCCT	TACTTCAAA	AGTCAATATG	CGATGGTGA	CACTGTCTGCC	4620
Db		4561	CTCTTTCTCT	TGCTTGCCT	TACTTCAAA	AGTCAATATG	CGATGGTGA	CACTGTCTGCC	4620
QY		4621	AAAGAGTTTGT	CAATATAAG	ATTCCTTTCT	TGTAACATTT	CTACCTTGG	GGGTTCAAT	4680
Db		4621	AAAGAGTTTGT	CAATATAAG	ATTCCTTTCT	TGTAACATTT	CTACCTTGG	GGGTTCAAT	4680
QY		4681	ATAATCAAA	AGGAGTACT	GTAACTGCA	AAAAAAGCT	ACCTGTG	ACATATATATGT	4740
Db		4681	ATAATCAAA	AGGAGTACT	GTAACTGCA	AAAAAAGCT	ACCTGTG	ACATATATATGT	4740
QY		4741	GATGGTTAC	CTGCAGTA	AGGTGGTCA	ATAATAAT	TAATACAG	AAATGAACCG	4800
Db		4741	GATGGTTAC	CTGCAGTA	AGGTGGTCA	ATAATAAT	TAATACAG	AAATGAACCG	4800
QY		4801	AGCAGAACTG	TCAGAGAAAT	TGTCAGAAAT	TCACACTCT	GAAACACG	GGCTATACAGTAA	4860
Db		4801	AGCAGAACTG	TCAGAGAAAT	TGTCAGAAAT	TCACACTCT	GAAACACG	GGCTATACAGTAA	4860
QY		4861	TAATCATAA	TAAATAGC	CACTCAAT	CAAAACAT	CACTGGCG	ACCTGTGCACATATAA	4920
Db		4861	TAATCATAA	TAAATAGC	CACTCAAT	CAAAACAT	CACTGGCG	ACCTGTGCACATATAA	4920
QY		4921	TCAGTGGAG	ATGTGATT	GAAGCACA	AGGCTTAA	GTGAATGCT	GTAGAGAGCTAATGTATTC	4980
Db		4921	TCAGTGGAG	ATGTGATT	GAAGCACA	AGGCTTAA	GTGAATGCT	GTAGAGAGCTAATGTATTC	4980
QY		4981	ATTTTATGAA	ATTTTAC	TATTTTAA	TGTGTATCC	CTGACCAT	CTTGAACTTTTACTT	5040
Db		4981	ATTTTATGAA	ATTTTAC	TATTTTAA	TGTGTATCC	CTGACCAT	CTTGAACTTTTACTT	5040
QY		5041	GAAGATTTAT	TTTTTTTT	TTTAAATCA	CTGTTTAT	TAGATTTAG	ATTTCTGCTTTGTT	5100
Db		5041	GAAGATTTAT	TTTTTTTT	TTTAAATCA	CTGTTTAT	TAGATTTAG	ATTTCTGCTTTGTT	5100
QY		5101	TTTCTTTTAT	CTATGATG	TTTTTTTAT	TTTTTTTAT	TGAGCTG	CTCTTAAAGCTTCATCAA	5160
Db		5101	TTTCTTTTAT	CTATGATG	TTTTTTTAT	TTTTTTTAT	TGAGCTG	CTCTTAAAGCTTCATCAA	5160
QY		5161	TGAGAGAAAT	GTATTAAT	TCCATTTAT	CTTTAC	CCCTAAAA	AAAAA	5217
Db		5161	TGAGAGAAAT	GTATTAAT	TCCATTTAT	CTTTAC	CCCTAAAA	AAAAA	5217

RESULT 2
AX467723
LOCUS

5092 bp DNA

linear . PAT 16-JUL-2002

D	b	2756	TTCCTCTACTACAGATTTTTCGAGAGGTAAATGTAATTGATCCAGAGGTGAGAACTCACTG	2815
Q	y	2941	CCTTTTATAATGCCGATTAAATTGCATGAACAAGTTTCCAACAATAAACAGTAAATAAAA	3000
D	b	2816	CCTTTTAAATGCCGATTAAATGCAATGAACAAGTTTCCAACAATAAACAGTAAATAAAA	2875
Q	y	3001	GAAACATGTATTAGCACTTAATAAGCCAGGGGCTGGACGACGTGTGTACATGCTTTCAA	3060
D	b	2876	GAAACATGTATTAGCACTTAATAAGCCAGGGGCTGGACGACGTGTGTACATGCTTTCAA	2935
Q	y	3061	TCCATGAACGTGTTAACTGTTAGTATCTCTATTGGACATGTGAGGAACCAATGGA	3120
D	b	2936	TCCATGAACGTGTTAACTGTTAGTATCTCTATTGGACATGTGAGGAACCAATGGA	2995
Q	y	3121	GTTCATAACAGTAGAGTTAAAAATTAATCTTCATATATTATATTGCTCAATCTCACAG	3180
D	b	2996	GTTCATAACAGTAGAGTTAAAAATTAATCTTCATATATTATATTGCTCAATCTCACAG	3055
Q	y	3181	ACATCTCTGCTACCAAAAGCTATCATATCTAGATATCGGCATAAAGATGACCTTGGGGC	3240
D	b	3056	ACATCTCTGCTACCAAAAGCTATCATATCTAGATATCGGCATAAAGATGACCTTGGGGC	3115
Q	y	3241	ACACTAGAAATCTTTGAGAGAAATCTGGCAGAGAAAACAATAATTTATCCCTACATAAA	3300
D	b	3116	ACACTAGAAATCTTTGAGAGAAATCTGGCAGAGAAAACAATAATTTATTCCTACATAAA	3175
Q	y	3301	ACCAGCATTTTACAGGTTTTATTTTAACTATGTAAGTATGTTATCTGTATCTTTCATA	3360
D	b	3176	ACCAGCATTTTACAGGTTTTATTTTAACTATGTAAGTATGTTATCTGTATCTTTCATA	3235
Q	y	3361	TAAGTGTCCCCGGAATTTATTTCTCTGTGGGTTCTTGTCTCGCTGACTCCCAAGAAATG	3420
D	b	3236	TAAGTGTCCCCGGAATTTATTTCTCTGTGGGTTCTTGTCTCGCTGACTCCCAAGAAATG	3295
Q	y	3421	AAACCCGAGACCTTTGAGGTGAGTGTACAGATTTCTTAAAGATGGTGTGTACAGTTTGT	3480
D	b	3296	AAACCCGAGACCTTTGAGGTGAGTGTACAGATTTCTTAAAGATGGTGTGTGTACAGTTTGT	3355
Q	y	3481	TCCTTCAGATGTTACAGATGTGTCGGAGTTTCTCCCTTATGTTGAGTTCGTGCTCTCGCT	3540
D	b	3356	TCCTTCAGATGTTACAGATGTGTCGGAGTTTCTCCCTTATGTTGAGTTCGTGCTCTCGCT	3415
Q	y	3541	GACTTCAACAAATGAAGCCGAGACCTTTGCAAGTGAAGTGTGACAGTTCTTAAAGGCAGT	3600
D	b	3416	GACTTCAACAAATGAAGCCGAGACCTTTGCAAGTGAAGTGTGACAGTTCTTAAAGGCAGT	3475
Q	y	3601	GCGTCCAGAGTTGTTGTTCTCCCGGTAGTGTGCTGGTTCGCTGATGTCAGGAATGAA	3660
D	b	3476	GCGTCCAGAGTTGTTGTTCTCCCGGTAGTGTGCTGGTTCGCTGATGFCAGGAATGAA	3535
Q	y	3661	GCTGCAGACCTCGCGTAAAGTGTACAGCTCATAAAGTGAAGTGCACAAACCAACAGTGA	3720
D	b	3536	GCTGCAGACCTCGCGTAAAGTGTACAGCTCATAAAGTGAAGTGCACAAACCAACAGTGA	3595
Q	y	3721	GCAGTAGCAAGATTTATTTAAGAGCAAAAGAACAAAGCTTCCCACCACATAGAAACGGA	3780
D	b	3596	GCAGTAGCAAGATTTATTTAAGAGCAAAAGAACAAAGCTTCCCACCACATAGAAACGGA	3655
Q	y	3781	CCAGAAATGTTGCTGCTGCTGGTAGCCAGCTTTTATTCCTTTATTTGGCCACACCCCA	3840
D	b	3656	CCAGAAATGTTGCTGCTGCTGGTAGCCAGCTTTTATTCCTTTATTTGGCCACACCCCA	3715
Q	y	3841	CATCCTCTGATTGGCCCAATTTACAGATGCTGATTGGTCCCAATTTTATAGCGTGCTGAT	3900
D	b	3716	CATCCTCTGATTGGCCCAATTTTACAGATGCTGATTGGTCCCAATTTTATAGCGTGCTGAT	3775
Q	y	3901	TGTTGCGTTTTTACAGAGTCTGATTGGTGCATTTTACAAATCCCTTTAGCTAGACACAGAGT	3960
D	b	3776	TGTTGCGTTTTTACAGAGTCTGATTGGTGCATTTTACAAATCCCTTTAGCTAGACACAGAGT	3835
Q	y	3961	GCTGATTGGTGCCTTTATAATCCCTTTAGCTAGACACAAAGTTCTTACAAGTCCCCACCCA	4020
D	b	3836	GCTGATTGGTGCCTTTATAATCCCTTTAGCTAGACACAAAGTTCTTACAAGTCCCCACCCA	3895

```

QY 5101 TTCTCTTTTATCTATGATGATTTTATTTTATGTCAGTGTCTTAAAGCTTCATCAA 5160
Db 4976 TTCTCTTTTATCTATGATGATTTTATTTTATGTCAGTGTCTTAAAGCTTCATCAA 5035
QY 5161 TGAGAAGAAATGATTAAATCCATTATTTCTTACCCCTAAAAAATTTTTTTTTT 5217
Db 5036 TGAGAAGAAATGATTAAATCCATTATTTCTTACCCCTAAAAAATTTTTTTTTT 5092

RESULT 3
AC013401
LOCUS Homo sapiens BAC clone RP11-98N11 from 2, complete sequence.
DEFINITION Homo sapiens BAC clone RP11-98N11 from 2, complete sequence.
ACCESSION AC013401
VERSION AC013401.10 GI:13435268
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 168595)
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
9847074
2 (bases 1 to 168595)
Sandbothe, T., Maupia, R. and Barrett, M.
The sequence of Homo sapiens BAC clone RP11-98N11
Unpublished
3 (bases 1 to 168595)
Waterston, R.H.
Direct Submission
Submitted (09-NOV-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 168595)
Waterston, R.H.
Direct Submission
Submitted (23-MAR-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 168595)
Waterston, R.
Direct Submission
Submitted (09-MAY-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Mar 23, 2001 this sequence version replaced gi:11192165.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
-----
Center project name: H_NH0098N11
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by

restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenoe, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-814A15. Actual start of this clone is at base position 1 of RP11-98N11; actual end is at base position 168595 of RP11-98N11.

The sequence RP11-98N11 from base position 1 to 165 is covered by a single plasmid subclone. The region is double-stranded and sequence quality is greater than phred 30 throughout.

FEATURES

source

Location/Qualifiers

1..168595

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="2"

/map="2"

/clone="RP11-98N11"

/clone_lib="RPCI-11"

3..118

/rpt_family="L1"

282..391

/rpt_family="L1"

397..881

/rpt_family="L1"

2044..2087

/rpt_family="(CA)n"

2716..2767

/rpt_family="MIR"

3357..3378

/rpt_family="AT-rich"

4191..4337

/rpt_family="MIR"

4361..4391

/rpt_family="(TG)n"

5304..5358

/rpt_family="L2"

5579..5607

/rpt_family="AT-rich"

5730..5752

/rpt_family="AT-rich"

5771..5850

/rpt_family="Alu"

6223..6526

/rpt_family="Alu"

6534..6678

/rpt_family="MIR"

6750..6875

/rpt_family="MIR"

7352..7496

/rpt_family="L1"

7615..7643

/rpt_family="AT-rich"

7982..8313

/rpt_family="L1"

8736..9031

/rpt_family="Alu"

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region	9202..9263	/rpt_family="AT_rich"	Qy	483	GAACATAAAACAGCTCCTGGAAGAGGACTATGACATCATCAGGTGGAGTCTCCAGGGA	542
repeat_region	9636..9707	/rpt_family="L1"	Db	132799	GAACATAAAACAGCTCCTGGAAGAGGACTATGACATCATCAGGTGGAGTCTCCAGGGA	132858
repeat_region	9708..9869	/rpt_family="Alu"	Qy	543	CAGCGGACCCCTTTGGAAAAGGACTAGAAAGTGTGAAATCTATTAGTCTTCGATATGAAT	602
repeat_region	9870..10375	/rpt_family="L1"	Db	132859	CAGCGGACCCCTTTGGAAAAGGACTAGAAAGTGTGAAATCTATTAGTCTTCGATATGAAT	132918
repeat_region	10375..10748	/rpt_family="L1"	Qy	603	TCTCTGTCTGTGTAAAGCATTTTCATATTACAAGACACAGGCGCTACTCTTAGGCGACGA	662
repeat_region	11052..11336	/rpt_family="Alu"	Db	132919	TCTCTGTCTGTGTAAAGCATTTTCATATTACAAGACACAGGCGCTACTCTTAGGCGACGA	132978
repeat_region	11733..11939	/rpt_family="L2"	Qy	663	AAAAGTGGCAACAGCGACAGAGGAGGAAAGAGATCATGAGGCATTTTCAGAGTGCACGT	722
repeat_region	12053..12183	/rpt_family="L2"	Db	132979	AAAAGTGGCAACAGCGACAGAGGAGGAAAGAGATCATGAGGCATTTTCAGAGTGCACGT	133038
repeat_region	12220..12770	/rpt_family="ERV1"	Qy	723	CTTTTCATATATTTCTCAATGCCGTATGTTTGGTGTATTTTGGCCCAAGCATAACAATCT	782
repeat_region	12778..12994	/rpt_family="L2"	Db	133039	CTTTTCATATATTTCTCAATGCCGTATGTTTGGTGTATTTTGGCCCAAGCATAACAATCT	133098
repeat_region	13029..13457	/rpt_family="ERV1"	Qy	783	GCTCAAGAAAAAATCTGGAGAAAAAAGGTGCGCTTTGCCCAATGTTTGTCTCTTTT	842
repeat_region	13461..13784	/rpt_family="L2"	Db	133099	GCTCAAGAAAAAATCTGGAGAAAAAAGGTGCGCTTTGCCCAATGTTTGTCTCTTTT	133158
repeat_region	14947..14994	/rpt_family="(CA)n"	Qy	843	TGACAAGCCCTGAGATTTCTGAGGGGATTCACATAAATGGGATCAGTCAATTCATTTAC	902
repeat_region	16734..16769	/rpt_family="(CA)n"	Db	133159	TGACAAGCCCTGAGATTTCTGAGGGGATTCACATAAATGGGATCAGTCAATTCATTTAC	133218
repeat_region	17262..17302	/rpt_family="L1"	Qy	903	GTTGTGTGCAAAATATGATTTAAAGATACAACTTTGCGAGAGAGCATGCTTTCCTAAGGGT	962
repeat_region	17696..18608	/rpt_family="ERVK"	Db	133219	GTTGTGTGCAAAATATGATTTAAAGATACAACTTTGCGAGAGAGCATGCTTTCCTAAGGGT	133278
repeat_region	18651..18683	/rpt_family="(TG)n"	Qy	963	AGGCACGTGGAGGACTAAGGGTAAAGCATTTCTCAAGATCAGTTAAATCAAGAAAGGTGCT	1022
repeat_region	20152..20271	/rpt_family="MIR"	Db	133279	AGGCACGTGGAGGACTAAGGGTAAAGCATTTCTCAAGATCAGTTAAATCAAGAAAGGTGCT	133338
repeat_region	20443..20836	/rpt_family="L2"	Qy	1023	CTTTGCATTTCTGAAATGCCCTTGTGCAAAATATTGGTTATATTGATTTAAATTTACACTTA	1082
repeat_region	21214..21271	/rpt_family="T-rich"	Db	133339	CTTTGCATTTCTGAAATGCCCTTGTGCAAAATATTGGTTATATTGATTTAAATTTACACTTA	133398
repeat_region	21731..21843	/rpt_family="MIR"	Qy	1083	ATGGAACAAACCTTTAACTTACAGATGAACAAACCCACAAAGCAAAATAATCAAAAGCCC	1142
repeat_region	24429..24451	/rpt_family="(TG)n"	Db	133399	ATGGAACAAACCTTTAACTTACAGATGAACAAACCCACAAAGCAAAATAATCAAAAGCCC	133458
repeat_region	24512..24551	/rpt_family="AT_rich"	Qy	1143	TACCTATGATTTCATATTTTCTGTGTAACCTGGATTAAGGATTCCTGCTGCTTTTGGGC	1202
repeat_region	24558..24712	/rpt_family="MIR"	Db	133459	TACCTATGATTTCATATTTTCTGTGTAACCTGGATTAAGGATTCCTGCTGCTTTTGGGC	133518
repeat_region	24838..25477	/rpt_family="MERL_type"	Qy	1203	ATAAATGATAATGGAATATTTCAGGTAATTTTAAATGAGGCGCCATCTACAAATCT	1262
repeat_region	26104..26655	/rpt_family="AT_rich"	Db	133519	ATAAATGATAATGGAATATTTCAGGTAATTTTAAATGAGGCGCCATCTACAAATCT	133578
repeat_region	26716..27433	/rpt_family="ERV1"	Qy	1263	TAGCAATACCTTTGGATAAATCTAAATTCAGCTGGACATGTCFAATTTGTTTTTATATA	1322
repeat_region	27659..27707	/rpt_family="L1"	Db	133579	TAGCAATACCTTTGGATAAATCTAAATTCAGCTGGACATGTCFAATTTGTTTTTATATA	133638
repeat_region	28264..28456	/rpt_family="L1"	Qy	1323	CATCTTTGCTAGAAATTTCAAAATTTTAAAGTATGTAATTTAGTTAATAGCTGTCTGATC	1382
repeat_region	28765..28855	/rpt_family="MERL_type"	Db	133639	CATCTTTGCTAGAAATTTCAAAATTTTAAAGTATGTAATTTAGTTAATAGCTGTCTGATC	133698
repeat_region	28988..29238	/rpt_family="L1"	Qy	1383	AATTCAAAAACATTTACTTTCTCTAAATTTTAAAGTATGTAATTTAGTTAATAGCTGTCTGATC	1442
misc_feature	29346..29820	/note="similar to EST BF589745 (NID:g11682069)"	Db	133699	AATTCAAAAACATTTACTTTCTCTAAATTTTAAAGTATGTAATTTAGTTAATAGCTGTCTGATC	133758
repeat_region	31395..31415		Qy	1443	ATCTACACATACAATTTATAGATTGTTTTTTCATTATATAATGTTTCATCTTAACAGAAATGT	1502
			Db	133759	ATCTACACATACAATTTATAGATTGTTTTTTCATTATATAATGTTTCATCTTAACAGAAATGT	133818
			Qy	1503	CTTTGTGATTGTTTTTAGAAAACTGAGAGTTTTTAATTCATAATTAATCTGATCAAAAAATT	1562
			Db	133819	CTTTGTGATTGTTTTTAGAAAACTGAGAGTTTTTAATTCATAATTAATCTGATCAAAAAATT	133878

Query Match 91.3%; Score 4763.6; DB 9; Length 168595;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 4775; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
Qy 423 AACTACTGGGATACAGCAGGCTTGGTGCATTTGACATTTGACATTCATGATATCAGCCAAAGTG 482
Db 132739 AAATATTTCTCACCCAGCAGGCTTGGTGCATTTGACATTTGACATTCATGATATCAGCCAAAGTG 132798

QY	1563	GTGGGAACAATCCAGCATTAATGTATGTGTGATGTGTTTTATGTACATAAGGAGTCTTAAAG	1622
DB	133879	GTGGGAACAATCCAGCATTAATGTATGTGTGATGTGTTTTATGTACATAAGGAGTCTTAAAG	133938
QY	1623	CTTGGTGCTTGAAGTCTTTTGTACTTATAGTCCCATGTTTAAATTTACTACTTTATATCTA	1682
DB	133939	CTTGGTGCTTGAAGTCTTTTGTACTTATAGTCCCATGTTTAAATTTACTACTTTATATCTA	133998
QY	1683	AAGCATTTATGTTTTTCCAATTCAAATTTACATGATGCTAAATTTATGGCAATTTAACAATA	1742
DB	133999	AAGCATTTATGTTTTTCCAATTCAAATTTACATGATGCTAAATTTATGGCAATTTAACAATA	134058
QY	1743	TTAAGATTTCCGAATAGAATATGTGAATTCCTTCACATACATAGAAATGAAAGTTCATT	1802
DB	134059	TTAAGATTTCCGAATAGAATATGTGAATTCCTTCACATACATAGAAATGAAAGTTCATT	134118
QY	1803	TCGTAAAGCAAGATGCTGGGTGAAGAGTGTCTTTTGATTGAAAGATCACTAGATTAGTAG	1862
DB	134119	TCGTAAAGCAAGATGCTGGGTGAAGAGTGTCTTTTGATTGAAAGATCACTAGATTAGTAG	134178
QY	1863	AGGCAAGACACTTCTAGTCCCTTAATCTACCCCTTAATAGCCATGTGGTCACGTGAAGTCAG	1922
DB	134179	AGGCAAGACACTTCTAGTCCCTTAATCTACCCCTTAATAGCCATGTGGTCACGTGAAGTCAG	134238
QY	1923	TGAACCCATCTATCTCCTCATCTATTTTTTTCATCTCTPAAATGAGGATATAAATTAAGC	1982
DB	134239	TGAACCCATCTATCTCCTCATCTATTTTTTTCATCTCTPAAATGAGGATATAAATTAAGC	134298
QY	1983	TCCTCATTTTTTTTTTTTTTTTGGATAGAGTTTTTGCCTTGTCCACCAGGTTGGAGTGCA	2042
DB	134299	TCCTCATTTTTTTTTTTTTTTTGGATAGAGTTTTTGCCTTGTCCACCAGGTTGGAGTGCA	134358
QY	2043	ATGGCAGCATCTCAGCTCACGTGCAACCCCTCTGCTTCTCTCGGTTCAAGTGATTCTCCTGCT	2102
DB	134359	ATGGCAGCATCTCAGCTCACGTGCAACCCCTCTGCTTCTCTCGGTTCAAGTGATTCTCCTGCT	134418
QY	2103	TCAGCTCCCAAGTAGCGGGATACAGGTGCCGCCACCACATCTGTGCTAATTTTTTGT	2162
DB	134419	TCAGCTCCCAAGTAGCGGGATACAGGTGCCGCCACCACATCTGTGCTAATTTTTTGT	134478
QY	2163	ATTTTCACCATGTTGGCCAGGCTGGTCTCGAACCCCTACCTTCAGGTGATCCCTGCCCTCG	2222
DB	134479	ATTTTCACCATGTTGGCCAGGCTGGTCTCGAACCCCTACCTTCAGGTGATCCCTGCCCTCG	134538
QY	2223	GCCTCTCAAAGTGCTGGGATTTACAGGTGTAGCCACCACACGCCGCCCAATATCAGTTTTT	2282
DB	134539	GCCTCTCAAAGTGCTGGGATTTACAGGTGTAGCCACCACACGCCGCCCAATATCAGTTTTT	134598
QY	2283	TCCTTTTTAACACAAGGCTAACACAATCAAAATACTAGCTAGGGGAGAAAAAAATAA	2342
DB	134599	TCCTTTTTAACACAAGGCTAACACAATCAAAATACTAGCTAGGGGAGAAAAAAATAA	134658
QY	2343	GGCACTGTTTATGTATACAGGCTCTGTTGCAATCACTGGCAGACAAATTAACAGTAG	2402
DB	134659	GGCACTGTTTATGTATACAGGCTCTGTTGCAATCACTGGCAGACAAATTAACAGTAG	134718
QY	2403	AATCAATCCTTTTCATATATCCTTCTGCAGAAATACATAAAATCCACAAATGGCTATCT	2462
DB	134719	AATCAATCCTTTTCATATATCCTTCTGCAGAAATACATAAAATCCACAAATGGCTATCT	134778
QY	2463	TCCTTTTATGATATTTGGGAAATGTAGTCAAGTGACAGATATTTTGGTGGGTGATA	2522
DB	134779	TCCTTTTATGATATTTGGGAAATGTAGTCAAGTGACAGATATTTTGGTGGGTGATA	134838
QY	2523	GACCACAAAGGACTGTGTTTGATGATGGTTTGCAATAAAATTTATACCTTAGTTTTTAC	2582
DB	134839	GACCACAAAGGACTGTGTTTGATGATGGTTTGCAATAAAATTTATACCTTAGTTTTTAC	134898
QY	2583	GTATGTTACATGTTTAGATTTAGAGTATGAAATTTAGTAGGAGGATTAATTAACAAGAAC	2642
DB	134899	GTATGTTACATGTTTAGATTTAGAGTATGAAATTTAGTAGGAGGATTAATTAACAAGAAC	134958
QY	2643	AGGCAAGAGGAGTAGAATTAACCCCTCTCTTAATACCTGTGCACAAGTAGGCTTTTCAGA	2702

Db 136039 AGTACGAGATTTATTATGAAGACAAAGAACAAAGCTTCCACCACATAGAAACGGACC 136098
Qy 3783 AGAATTGGTCTGCTGCTGCTGAGCCAGCTTTTATTCCTCTTATTTGGCCACACCCACA 3842
Db 136099 AGAATTGGTCTGCTGCTGCTGAGCCAGCTTTTATTCCTCTTATTTGGCCACACCCACA 136158
Qy 3843 TCCTGCTGATTTGGCCCAATTTTACAGAATGCTGATTTGGTCCCAATTTATAGCGTCTGATTTG 3902
Db 136159 TCCTGCTGATTTGGCCCAATTTTACAGAATGCTGATTTGGTCCCAATTTATAGCGTCTGATTTG 136218
Qy 3903 GTGGCTTTTACAGAGTCTGATTTGGTGCAATTTTACAAATCCTTTAGCTAGACACAGATGC 3962
Db 136219 GTGGCTTTTACAGAGTCTGATTTGGTGCAATTTTACAAATCCTTTAGCTAGACACAGATGC 136278
Qy 3963 TGATTGGTGGCTTTATATCTTTAGCTAGACACAAAGTCTCAAGTCCCCACCCAAAC 4022
Db 136279 TGATTGGTGGCTTTATATCTTTAGCTAGACACAAAGTCTCAAGTCCCCACCCAAAC 136338
Qy 4023 CCAGAAGCTCGGCTGGCTTCACTCTCTGTAAGGAAATTTAGGTTTCAACAAAGTTTCAAAG 4082
Db 136339 CCAGAAGCTCGGCTGGCTTCACTCTCTGTAAGGAAATTTAGGTTTCAACAAAGTTTCAAAG 136398
Qy 4083 TGCTAAACACTACAGTTTCTCATCTCTGCAACTCGATTTCCACTCATGTGTTGAATCCC 4142
Db 136399 TGCTAAACACTACAGTTTCTCATCTCTGCAACTCGATTTCCACTCATGTGTTGAATCCC 136458
Qy 4143 AGGCTCTAAGACTTAACTTGCATCTGTGACTTTATGTTCTCTCAATTTACACAAAGCT 4202
Db 136459 AGGCTCTAAGACTTAACTTGCATCTGTGACTTTATGTTCTCTCAATTTACACAAAGCT 136518
Qy 4203 ACTATCTGTACATCTCTGCTGTTAACTTCCAGACTAACTTCTTTGATTTCACAAATGAC 4262
Db 136519 ACTATCTGTACATCTCTGCTGTTAACTTCCAGACTAACTTCTTTGATTTCACAAATGAC 136578
Qy 4263 CACACACTTTTGGTGGCTTTGCTATCGGTTTATGTTACTGTTTAAATAGAGAGCTTC 4322
Db 136579 CACACACTTTTGGTGGCTTTGCTATCGGTTTATGTTACTGTTTAAATAGAGAGCTTC 136638
Qy 4323 TTCAGAAATTTAGTAGAGGAGAGAGAGAGTACACATCTCTAAATGTACCATGCCT 4382
Db 136639 TTCAGAAATTTAGTAGAGGAGAGAGAGTACACATCTCTAAATGTACCATGCCT 136698
Qy 4383 TTCAGTCAAGCATCCCTATCAGTGGCTGTCAAGGGTGGCTCAGAAATAGGTAGAGTT 4442
Db 136699 TTCAGTCAAGCATCCCTATCAGTGGCTGTCAAGGGTGGCTCAGAAATAGGTAGAGTT 136758
Qy 4443 AAGNATTAAGTAAATTTGTTAGGATGAAGGATTCATCTTAAAGCTTATATTACAT 4502
Db 136759 AAGNATTAAGTAAATTTGTTAGGATGAAGGATTCATCTTAAAGCTTATATTACAT 136818
Qy 4503 CAAGTAAATTTAAATTAATTTGAACATTTTCCAGGCATCCCTGTTATTTATTTGCTCT 4562
Db 136819 CAAGTAAATTTAAATTAATTTGAACATTTTCCAGGCATCCCTGTTATTTATTTGCTCT 136878
Qy 4563 CTTTCTGCTGCTACTTCAAAAGTCATATGSCATGGTGTGACTAGAACTGTCTCGCCAA 4622
Db 136879 CTTTCTGCTGCTACTTCAAAAGTCATATGSCATGGTGTGACTAGAACTGTCTCGCCAA 136938
Qy 4623 AGAGTTTGTCAATATAGATTCCTTTTGTAAACATTTCTACCTGGGGCTTCATTTAT 4682
Db 136939 AGAGTTTGTCAATATAGATTCCTTTTGTAAACATTTCTACCTGGGGCTTCATTTAT 136998
Qy 4683 AATCAAAAGGAGTACTGTAACCTGTCAAAAAGGCTTACCTGTGACAAATATATTGTGA 4742
Db 136999 AATCAAAAGGAGTACTGTAACCTGTCAAAAAGGCTTACCTGTGACAAATATATTGTGA 137058
Qy 4743 TGTTTACCTGCTAGTGGTGGCAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 4802
Db 137059 TGTTTACCTGCTAGTGGTGGCAATAATAATAATAATAATAATAATAATAATAATAATAATA 137118
Qy 4803 CAGAACTGTGAGAAATGCTGAGATTTACACTCTGGAAGAACACGGCTATACAGTAATA 4862
Db 137119 CAGAACTGTGAGAAATGCTGAGATTTACACTCTGGAAGAACACGGCTATACAGTAATA 137178

Qy 4863 ATCATAATAATAGCCACTCAATCAAAAACATCACTGGCGACTTGTCCACATATATAATC 4922
Db 137179 ATCATAATAATAGCCACTCAATCAAAAACATCACTGGCGACTTGTCCACATATATAATC 137238
Qy 4923 AGTGAGATGTGATTGAAGACAAAGGCTTAAAGTGAATGCTAGAGAGCTAAATGATCAT 4982
Db 137239 AGTGAGATGTGATTGAAGACAAAGGCTTAAAGTGAATGCTAGAGAGCTAAATGATCAT 137298
Qy 4983 TTTTATGGAATTTTACTTATTTAAATGTCATCCCTGACCATCTTGAACCTTTTACTTGA 5042
Db 137299 TTTTATGGAATTTTACTTATTTAAATGTCATCCCTGACCATCTTGAACCTTTTACTTGA 137358
Qy 5043 AGATTTATTTTATTTTAAATCACTGTTTATTTAGATTTAGTATTTCTGGTCTTGTGTTT 5102
Db 137359 AGATTTATTTTATTTTAAATCACTGTTTATTTAGATTTAGTATTTCTGGTCTTGTGTTT 137418
Qy 5103 TCTTTTATCTATGATGATTTTATTTTATTTTATGAGTCTCTTAAAGTTCATCAATG 5162
Db 137419 TCTTTTATCTATGATGATTTTATTTTATTTTATGAGTCTCTTAAAGTTCATCAATG 137478
Qy 5163 AGAAGAAATGTTTAAATCCATTTATTTTACCTTAAATTTTAAAAA 5216
Db 137479 AGAAGAAATGTTTAAATCCATTTATTTTACCTTAAATTTTAAAAA 137532

RESULT 4
AX467725
LOCUS AX467725 32463 bp DNA linear PAT 16-JUL-2002
DEFINITION Sequence 5 from Patent WO0244332.
ACCESSION AX467725
VERSION AX467725.1 GI:21900897
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 Sun, Z., Li, X. and Jay, G.
AUTHORS Prostate polynucleotides and uses
TITLE Patent: WO 0244332-A 5 06-JUN-2002;
JOURNAL Origene Technologies Inc. (US)
FEATURES Location/Qualifiers
source 1. 32463
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

BASE COUNT 10196 a 5729 c 5685 g 10853 t
ORIGIN

Query Match 91.28; Score 4758.4; DB 6; Length 32463;
Best Local Similarity 99.88; Pred. No. 0;
Matches 4765; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 423 AAGTAACCTGGGATTTACAGAGGCTTGGTGGATTTGACATTTGACATTCATGATATCAGCCAAAGTG 482
Db 27688 AATATTTCTCACCCAGAGGCTTGGTGGATTTGACATTCATGATATCAGCCAAAGTG 27747
Qy 483 GAACATAAAACAGCTCCTCGAAGAGGACTATGACATCATCAGGTTGGGAGTCTCCAGGGA 542
Db 27748 GAACATAAAACAGCTCCTCGAAGAGGACTATGACATCATCAGGTTGGGAGTCTCCAGGGA 27807
Qy 543 CAGCGGACCTTTGGAAAAGGACTAGAAAGTGAATATTTAGTCTTCGATATGAAT 602
Db 27808 CAGCGGACCTTTGGAAAAGGACTAGAAAGTGAATATTTAGTCTTCGATATGAAT 27867
Qy 603 TCTCTGCTCTGTAAAAAGCATTTTCAATTTTACAGACACAGGCTTCTCTAGGCGCAGCA 662
Db 27868 TCTCTGCTCTGTAAAAAGCATTTTCAATTTTACAGACACAGGCTTCTCTAGGCGCAGCA 27927
Qy 663 AAAAGTGGCAACAGGCAAGCAGAGGGAAGAAAGATCATGAGGCATTTTCAGAGTGCATGT 722
Db 27928 AAAAGTGGCAACAGGCAAGCAGAGGGAAGAAAGATCATGAGGCATTTTCAGAGTGCATGT 27987

this is the
Source
Substr
etal, #2

QY 723 CTTTTCATATATTTCTCAATGCCCTATGTTTGGTTTTATTTTGGCCCAAGCATACAATCT 782
DB 27988 CTTTTCATATATTTCTCAATGCCCTATGTTTGGTTTTATTTTGGCCCAAGCATACAATCT 28047
QY 783 GCTCAAGAAAAAATCTGGAGAAAAACAAGTGCCTTTTGGCCCAATGTTATGTTTCTTTT 842
DB 28048 GCTCAAGAAAAAATCTGGAGAAAAACAAGTGCCTTTTGGCCCAATGTTATGTTTCTTTT 28107
QY 843 TGACAAGCCCTGAGATTTCTGAGGGGAATTCACATAAATGGGATCAGGTCAATTCATTTAC 902
DB 28108 TGACAAGCCCTGAGATTTCTGAGGGGAATTCACATAAATGGGATCAGGTCAATTCATTTAC 28167
QY 903 GTTGTGTGCAAAATGATGTTTAAAGATACAACCTTTGAGAGAGCATGCTTTCCCTAAAGGT 962
DB 28168 GTTGTGTGCAAAATGATGTTTAAAGATACAACCTTTGAGAGAGCATGCTTTCCCTAAAGGT 28227
QY 963 AGGCACGTGGAGGACTAAGGGTAAAGCATTTCTCAAGATCAGTTAAATCAAGAAAGGTGCT 1022
DB 28228 AGGCACGTGGAGGACTAAGGGTAAAGCATTTCTCAAGATCAGTTAAATCAAGAAAGGTGCT 28287
QY 1023 CTTTGCATTTCTGAATGCCCTTGTGCAATATTTGGTTATATTCATTAATTTTACACTTA 1082
DB 28288 CTTTGCATTTCTGAATGCCCTTGTGCAATATTTGGTTATATTCATTAATTTTACACTTA 28347
QY 1083 ATGGAACAACCTTTTAACTTACAGATGAACAAACCCACAAAAGCAAAAATCAAAAGCCC 1142
DB 28348 ATGGAACAACCTTTTAACTTACAGATGAACAAACCCACAAAAGCAAAAATCAAAAGCCC 28407
QY 1143 TACCTATGATTTTCATATTTCTGTGTAACTGGATTAAGGATTCCTGCTGTTTGGGC 1202
DB 28408 TACCTATGATTTTCATATTTCTGTGTAACTGGATTAAGGATTCCTGCTGTTTGGGC 28467
QY 1203 ATAAATGATGAATTTTCCAGGTATGTTTAAATGAGGGCCCATCTACAAATCT 1262
DB 28468 ATAAATGATGAATTTTCCAGGTATGTTTAAATGAGGGCCCATCTACAAATCT 28527
QY 1263 TAGCAATACTTTGGATTAATTTCTAAATTCAGCTGCACATGCTCTAAATGTTTATATATA 1322
DB 28528 TAGCAATACTTTGGATTAATTTCTAAATTCAGCTGCACATGCTCTAAATGTTTATATA 28587
QY 1323 CATCTTGTGATGAATTTCAAAATTTAAATAGTATGTAATTTAGTTAATAGCTGCTGATC 1382
DB 28588 CATCTTGTGATGAATTTCAAAATTTAAATAGTATGTAATTTAGTTAATAGCTGCTGATC 28647
QY 1383 AATTCAAAACATTTACTTTCTTAAATTTAGACTATGAAGTCAATAAATTCACAAATAT 1442
DB 28648 AATTCAAAACATTTACTTTCTTAAATTTAGACTATGAAGTCAATAAATTCACAAATAT 28707
QY 1443 ATCTACACATACAATTTATAGATTTGTTTTCATTAATGTTCTTCATCTTAACAGAAATGT 1502
DB 28708 ATCTACACATACAATTTATAGATTTGTTTTCATTAATGTTCTTCATCTTAACAGAAATGT 28767
QY 1503 CTTTGTGATGTTTTAGAAAACCTGAGAGTTTTAAATTCATAATTTACTTGATCAAAAAAT 1562
DB 28768 CTTTGTGATGTTTTAGAAAACCTGAGAGTTTTAAATTCATAATTTACTTGATCAAAAAAT 28827
QY 1563 GTGGGAACATCCAGCATTAATTTGATGTAATTTGTTTATGTCATAGAAGGCTTTAAG 1622
DB 28828 GTGGGAACATCCAGCATTAATTTGATGTAATTTGTTTATGTCATAGAAGGCTTTAAG 28887
QY 1623 CTTGCTGCTTGAAGTCTTTTGTACTTAGTCCCATGTTTAAATTTACTTTTATATCTA 1682
DB 28888 CTTGCTGCTTGAAGTCTTTTGTACTTAGTCCCATGTTTAAATTTACTTTTATATCTA 28947
QY 1683 AAGCATTTATGTTTTTCAATTTCAATTTTACATGATGCTAAATTTATGGCAATTTATAACAATA 1742
DB 28948 AAGCATTTATGTTTTTCAATTTCAATTTTACATGATGCTAAATTTATGGCAATTTATAACAATA 29007
QY 1743 TTAAGATTTTGGAAATAGAATATGTAATTTTCCACATACATAGAAATGAAAGTTCATT 1802
DB 29008 TTAAGATTTTGGAAATAGAATATGTAATTTTCCACATACATAGAAATGAAAGTTCATT 29067

QY 1803 TCGTAAAGCAAGATGCTGGGTGAAGAGTGCTTTTGTATGGAAGATCACTAGATTAGTAG 1862
DB 29068 TCGTAAAGCAAGATGCTGGGTGAAGAGTGCTTTTGTATGGAAGATCACTAGATTAGTAG 29127
QY 1863 AGGCAAGACATCTAGTCCCTTAATCTACCCCTTAATAGCCATGTGGTCAAGTGAAGTCAG 1922
DB 29128 AGGCAAGACATCTAGTCCCTTAATCTACCCCTTAATAGCCATGTGGTCAAGTGAAGTCAG 29187
QY 1923 TGAACCCATCTCATCTCCCTCATACTTTTTTTCATCTCTAAATCAGGGTATAAATTTAAGC 1982
DB 29188 TGAACCCATCTCATCTCCCTCATACTTTTTTTCATCTCTAAATCAGGGTATAAATTTAAGC 29247
QY 1983 TCTTCATTTTTTTTTTTTTTTTGGATAGAGTTTTGCTCTCTGTCACCCAGGTGGAGTGCA 2042
DB 29248 TCTTCATTTTTTTTTTTTTTTTGGATAGAGTTTTGCTCTCTGTCACCCAGGTGGAGTGCA 29307
QY 2043 ATGCGACGATCTCAGCTCAGTCACTGCAACCCCTGCTCTCCGTTTCCAGTGATTCCTCGCT 2102
DB 29308 ATGCGACGATCTCAGCTCAGTCACTGCAACCCCTGCTCTCCGTTTCCAGTGATTCCTCGCT 29367
QY 2103 TCAGCCTCCCAAGTAGCCGGATTCACAGTGCCTCCGCCACACACATCTGCTCAATTTTGT 2162
DB 29368 TCAGCCTCCCAAGTAGCCGGATTCACAGTGCCTCCGCCACACACATCTGCTCAATTTTGT 29427
QY 2163 ATTTTCACCATGTTGGCCAGGCTGCTCGAACCCTTACCTCAGGTGATCCCTCGCCTCG 2222
DB 29428 ATTTTCACCATGTTGGCCAGGCTGCTCGAACCCTTACCTCAGGTGATCCCTCGCCTCG 29487
QY 2223 GCCTCTCAAAGTGTGCGGATTCACAGTGTGAGCCACACGCCCAAGCCCAATATCAGTTTT 2282
DB 29488 GCCTCTCAAAGTGTGCGGATTCACAGTGTGAGCCACACGCCCAAGCCCAATATCAGTTTT 29547
QY 2283 TCTTTTTTAAACAAAGCTTAACACAATCAAAATCTAGCTAGGGAGAAAAAATAA 2342
DB 29548 TCTTTTTTAAACAAAGCTTAACACAATCAAAATCTAGCTAGGGAGAAAAAATAA 29607
QY 2343 GGCACTGTTTATGTTAAACAGGCTTGTGTCATCTACTGGCAGACAATAAACAGTAAG 2402
DB 29608 GGCACTGTTTATGTTAAACAGGCTTGTGTCATCTACTGGCAGACAATAAACAGTAAG 29667
QY 2403 AATCAATCTCTTTTCATATATCTCTTGTGAGATACATAAATCCCAAAATGGCTATCT 2462
DB 29668 AATCAATCTCTTTTCATATATCTCTTGTGAGATACATAAATCCCAAAATGGCTATCT 29727
QY 2463 TCCTTTTTATCATATTTGGAGAAATGTAGCTTAAGTGCACAGATATTTTCTGGGTGATATA 2522
DB 29728 TCCTTTTTATCATATTTGGAGAAATGTAGCTTAAGTGCACAGATATTTTCTGGGTGATATA 29787
QY 2523 GACCACAAAGGACCTGTTTTCATGATGATGTTTTCATATAAAATTTATACCTTAGTTTTACTTT 2582
DB 29788 GACCACAAAGGACCTGTTTTCATGATGATGTTTTCATATAAAATTTATACCTTAGTTTTACTTT 29847
QY 2583 GTATGTTACATGTTTAGATTTAGATATGAATAATTTAGTAGGGAGGATTTAACAAGAAC 2642
DB 29848 GTATGTTACATGTTTAGATTTAGATATGAATAATTTAGTAGGGAGGATTTAACAAGAAC 29907
QY 2643 AGGCAAGAGAGTAGAATTAACCTCTCTTAATACCTGTGCACAAGTAGCTTTTCAGA 2702
DB 29908 AGGCAAGAGAGTAGAATTAACCTCTCTTAATACCTGTGCACAAGTAGCTTTTCAGA 29967
QY 2703 AACCTCAACCCCTACATAAATCTGATAGTTAGAAAAGCACACCTCCCAAGGAAGCGGTT 2762
DB 29968 AACCTCAACCCCTACATAAATCTGATAGTTAGAAAAGCACACCTCCCAAGGAAGCGGTT 30027
QY 2763 ATGTTTTGCAGTTTTGAATCAGAGAATAGAGCTATAGCAATCTTCATTTCTATAGTAACAT 2822
DB 30028 ATGTTTTGCAGTTTTGAATCAGAGAATAGAGCTATAGCAATCTTCATTTCTATAGTAACAT 30087
QY 2823 TAAAGAGCCTGGTTTTATTTATAGCAGTCATTAAGATTTAAAAATTTTACATCTTGGCGGTT 2882
DB 30088 TAAAGAGCCTGGTTTTATTTATAGCAGTCATTAAGATTTAAAAATTTTACATCTTGGCGGTT 30147
QY 2883 CTTTCTTACTCACATTTTCGAGAGGTAATGTAATGATCCACGAGGTGAGAAATCACTGCC 2942

|||||
Db 30148 CTTCTACTCACAGATTTTCGAGAGGTAATGTAATGATCCACGAGGTGAGAATCACTGCC 30207
QY 2943 TTTTATAATCGGATTAATTTGCATGACAAAGTTTCCAAACAATAACAGTAATAAANAAGA 3002
Db 30208 TTTTATAATCGGATTAATTTGCATGACAAAGTTTCCAAACAATAACAGTAATAAANAAGA 30267
QY 3003 AACATGTATTAGCACATTAATAAGCCAGGGCTGGACGACGTGTGTACATGCTTTCAATC 3062
Db 30268 AACATGTATTAGCACATTAATAAGCCAGGGCTGTACGACGTGTGTACATGCTTTCAATC 30327
QY 3063 CATGAACGTGTAACCTGGTACTAGTATCTCTATTGGACATGTGAGGAACCAATGGAGT 3122
Db 30328 CATGAACGTGTAACCTGGTACTAGTATCTCTATTGGACATGTGAGGAACCAATGGAGT 30387
QY 3123 TGATAAACAGTAGAGTTAAATAATCTCTTCATATATATTATGTCCTCAATCTCACAGAC 3182
Db 30388 TGATAAACAGTAGAGTTAAATAATCTCTTCATATATATTATGTCCTCAATCTCACAGAC 30447
QY 3183 ATCTCTGCTACCAAAAGCTATCATATCTAGATATGGGGCATAAGGATGACCTTGGGGCAC 3242
Db 30448 ATCTCTGCTACCAAAAGCTATCATATCTAGATATGGGGCATAAGGATGACCTTGGGGCAC 30507
QY 3243 ACTAGAAATCTTTGAGAGAAATCTGGCAGAGAAACAATAATTTATTCCTPACAATAAAAC 3302
Db 30508 ACTAGAAATCTTTGAGAGAAATCTGGCAGAGAAACAATAATTTATTCCTPACAATAAAAC 30567
QY 3303 CCAGCATTTTACAGGTTTATTTTAACTATGAGTATGTTATCTCTATCTCTTCATATA 3362
Db 30568 CCAGCATTTTACAGGTTTATTTTAACTATGAGTATGTTATCTCTATCTCTTCATATA 30627
QY 3363 AGTGTGCCCGGAATTTATTTCTCTGGTGGGTTCTTGGTCTCGCTGACCTCCAAGAATGAA 3422
Db 30628 AGTGTGCCCGGAATTTATTTCTCTGGTGGGTTCTTGGTCTCGCTGACCTCCAAGAATGAA 30687
QY 3423 ACCGACAGACCTTGAGGTGAGTGCACAGTTCCTTAAAGATGGTGTTCAGAGTTGTTC 3482
Db 30688 ACCGACAGACCTTGAGGTGAGTGCACAGTTCCTTAAAGATGGTGTTCAGAGTTGTTC 30747
QY 3483 CTTTCAGATGTTACAGTGTGTCGGAGTTCTCCCTTATGTTGAGTTCGTGGTCTCGCTGA 3542
Db 30748 CTTTCAGATGTTACAGTGTGTCGGAGTTCTCCCTTATGTTGAGTTCGTGGTCTCGCTGA 30807
QY 3543 CTTTCAACAATGAAGCCGACAGACCTTTGCAGTGAAGTGTGTGACAGTTCTTTAAAGCGAGTGC 3602
Db 30808 CTTTCAACAATGAAGCCGACAGACCTTTGCAGTGAAGTGTGTGACAGTTCTTTAAAGCGAGTGC 30867
QY 3603 GTCCAGATGTTGTTTCTCCCGGTAGGTTCTGTGTCTCGCTGTGATGTGAGGAATGAAGC 3662
Db 30868 GTCCAGATGTTGTTTCTCCCGGTAGGTTCTGTGTCTCGCTGTGATGTGAGGAATGAAGC 30927
QY 3663 TGCAGACCTTCGGGTAAGTGTTCACAGCTCATAAAGGTAGTGCAAAACCCAAACAGTGAGC 3722
Db 30928 TGCAGACCTTCGGGTAAGTGTTCACAGCTCATAAAGGTAGTGCAAAACCCAAACAGTGAGC 30987
QY 3723 AGTAGCAAGATTTATTTATGAAGAGCAAAAGAAAGCTTCCCCACCATAGAAACGGACC 3782
Db 30988 AGTAGCAAGATTTATTTATGAAGAGCAAAAGAAAGCTTCCCCACCATAGAAACGGACC 31047
QY 3783 AGAATTTGGTGTGCTGTGGTAGCCAGCTTTTATTCCTTATTTGGCCACACCCACA 3842
Db 31048 AGAATTTGGTGTGCTGTGGTAGCCAGCTTTTATTCCTTATTTGGCCACACCCACA 31107
QY 3843 TCCCTGCTGATTTGGCCATTTTACAGAAATGCTGATTTGTCATTTTATAGCGTGTGATTTG 3902
Db 31108 TCCCTGCTGATTTGGCCATTTTACAGAAATGCTGATTTGTCATTTTATAGCGTGTGATTTG 31167
QY 3903 GTGCGTTTTTACAGAGTGTGATTTGTTGCAATTTTACAAATCCTTTAGCTAGACACAGAGTGC 3962
Db 31168 GTGCGTTTTTACAGAGTGTGATTTGTTGCAATTTTACAAATCCTTTAGCTAGACACAGAGTGC 31227
QY 3963 TGATTTGGTGCCTTTATAATCCCTTTAGCTAGACAAAGTTCTTACAAAGTCCCCACCCCAAC 4022
|||||

Db 31228 TGATTTGGTGCCTTTATAATCCTTTAGCTAGACACAAAGTTCTACAAAGTCCCCACCCCAAC 31287
QY 4023 CCAGAAGCTCCGCTGGCTTACACCTCTCGTAAGGAATTTGAGTTTCAACAAAGTTTCAAG 4082
Db 31288 CCAGAAGCTCCGCTGGCTTACACCTCTCGTAAGGAATTTGAGTTTCAACAAAGTTTCAAG 31347
QY 4083 TGTCTAAACCTACAGTTTCTCATTTCTGTCAACCTGGATTTCCACTCATGTGTGTAATCCC 4142
Db 31348 TGTCTAAACCTACAGTTTCTCATTTCTGTCAACCTGGATTTCCACTCATGTGTGTAATCCC 31407
QY 4143 AGGCTCTAAGACTTAACTTGCCTATCTGTGACCTTTATGTCTCTGCAATTTACAAAGCT 4202
Db 31408 AGGCTCTAAGACTTAACTTGCCTATCTGTGACCTTTATGTCTCTGCAATTTACAAAGCT 31467
QY 4203 ACTATCTGTACACATCTCTGGTGTAACTTCAGACTTAACTTCTTTTGTATTCACAAATGAC 4262
Db 31468 ACTATCTGTACACATCTCTGGTGTAACTTCAGACTTAACTTCTTTTGTATTCACAAATGAC 31527
QY 4263 CACACACTTTTGGTTGAGGTTTGTCTATCGGTTTATTTGACTGGTTTAAATAGAGAGCTTC 4322
Db 31528 CACACACTTTTGGTTGAGGTTTGTCTATCGGTTTATTTGACTGGTTTAAATAGAGAGCTTC 31587
QY 4323 TTCCAGAAATTTGAGTAGATGGAGAGGAAATAGACATTTCTTAAAAATGTACATGCCT 4382
Db 31588 TTCCAGAAATTTGAGTAGATGGAGAGGAAATAGACATTTCTTAAAAATGTACATGCCT 31647
QY 4383 TTCAAGTTCACAAGCATCCCTTATCACATGGCTGTCAAGGGTGGCTCAGAATAGTAGAGTT 4442
Db 31648 TTCAAGTTCACAAGCATCCCTTATCACATGGCTGTCAAGGGTGGCTCAGAATAGTAGAGTT 31707
QY 4443 AAGAAATTTAAAGTAAATTTGGTGTAAAGCTTTCATCTTAAAGCTTATATTTACAT 4502
Db 31708 AAGAAATTTAAAGTAAATTTGGTGTAAAGCTTTCATCTTAAAGCTTATATTTACAT 31767
QY 4503 CAACCTGAATGTAAATTAATTTGGAACATTTTCCAGGATCCCTGTTATTTATTTGTCTCT 4562
Db 31768 CAACCTGAATGTAAATTAATTTGGAACATTTTCCAGGATCCCTGTTATTTATTTGTCTCT 31827
QY 4563 CTTTCCCTTGGCTTGCCTTACTTCAAAAGTTCATATGGCATGGTGTAGAACCTGTCTGCCAA 4622
Db 31828 CTTTCCCTTGGCTTGCCTTACTTCAAAAGTTCATATGGCATGGTGTAGAACCTGTCTGCCAA 31887
QY 4623 AGAGTTTGTCAATATAAGATTCCTTCTTTGTAAACATTTCTACCTTTGGGGCTTCATTTAT 4682
Db 31888 AGAGTTTGTCAATATAAGATTCCTTCTTTGTAAACATTTCTACCTTTGGGGCTTCATTTAT 31947
QY 4683 AATCAAAAGGAGTACTGTAACTGTCAAAAAGGCTACCTGTGACAAATATATTTATGTA 4742
Db 31948 AATCAAAAGGAGTACTGTAACTGTCAAAAAGGCTACCTGTGACAAATATATTTATGTA 32007
QY 4743 TGGTTACCTGACAGTAAGGTTGGGCAATAAATAAATAATATCAACAGAAATGAACCGAG 4802
Db 32008 TGGTTACCTGACAGTAAGGTTGGGCAATAAATAAATAATATCAACAGAAATGAACCGAG 32067
QY 4803 CAGAACTGTCAGAGAAATGTCAGAAATTCACACTCTGAAAGAACACCGCTTATACAGTAATA 4862
Db 32068 CAGAACTGTCAGAGAAATGTCAGAAATTCACACTCTGAAAGAACACCGCTTATACAGTAATA 32127
QY 4863 ATCATATAAATAGCCACTCAATCCAAAACATCACTGGGGACTTGTACATATATAATC 4922
Db 32128 ATCATATAAATAGCCACTCAATCCAAAACATCACTGGGGACTTGTACATATATAATC 32187
QY 4923 AGTGGAGATGTGATTGAAGCACAAGGCTTAAAGTGAATGTCTAGAGAGCTAAATGATTCAT 4982
Db 32188 AGTGGAGATGTGATTGAAGCACAAGGCTTAAAGTGAATGTCTAGAGAGCTAAATGATTCAT 32247
QY 4983 TTTTATGGAATTTTACTTTATTTTAAATGTCACTCCCTGACCACTTTTGAACCTTTTACTTGA 5042
Db 32248 TTTTATGGAATTTTACTTTATTTTAAATGTCACTCCCTGACCACTTTTGAACCTTTTACTTGA 32307
QY 5043 AGATTTATTTTATTTTAAATCACTGTTTATAGATTTAGTATTTCTGCTCTCTCTTTT 5102
Db 32308 AGATTTATTTTATTTTAAATCACTGTTTATAGATTTAGGTTATTTCTGCTCTCTCTTTT 32367

```

QY 5103 TCTTTTATCTATGATGATTTATTTATTTTATGTCAGTGCCCTTAAGCTTTCATCAATG 5162
Db 32368 TCTTTTATCTATGATGATTTATTTTATGTCAGTGCCCTTAAGCTTTCATCAATG 32427
QY 5163 AGAAGAAATGATTAATAATCCATTTATTTTACCCCT 5198
Db 32428 AGAAGAAATGATTAATAATCCATTTATTTTACCCCT 32463

RESULT 5
AF223389
LOCUS Homo sapiens PCGEM1 gene, non-coding mRNA. linear PRI 01-DEC-2000
DEFINITION AF223389
ACCESSION AF223389
VERSION AF223389.1 GI:11066459
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
REFERENCE
AUTHORS Srikantan,V., Zou,Z., Petrovics,G., Xu,L., Augustus,M., Davis,L.,
Livezey,J.R., Connell,T., Sesterhenn,I.A., Yoshino,K., Buzard,G.S.,
Mostofi,F.K., McLeod,D.G., Moul,J.W. and Srivastava,S.
TITLE PCGEM1, a prostate-specific gene, is overexpressed in prostate
cancer.
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (22), 12216-12221 (2000)
MEDLINE 20504490
PUBMED 11050243
REFERENCE
AUTHORS Srikantan,V., Zou,Z., Xu,L., Petrovics,G., Augustus,M., Davis,L.,
Livezey,J.R., Connell,T., Sesterhenn,I.A., Yoshino,K., Buzard,G.S.,
Mostofi,F.K., McLeod,D.G., Moul,J.W. and Srivastava,S.
Direct Submission
TITLE Submitted (11-JAN-2000) Center for Prostate Disease Research, 1530
JOURNAL East Jefferson Street, Rockville, MD 20852, USA
FEATURES
source
1. 1603
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="2q32"
<1..>1603
/gene="PCGEM1"
/note="detectable only in prostate; expression regulated
by androgen; has tumor-associated over-expression"
BASE COUNT 531 a 257 c 294 g 521 t
ORIGIN

Query Match 28.0%; Score 1458.4; DB 9; Length 1603;
Best Local Similarity 92.8%; Pred. No. 1.le-251;
Matches 1591; Conservative 0; Mismatches 1; Indels 122; Gaps 1;

QY 51 AAGGACATCTGCACCCAGTTTGGAACTGCAGTTTAAAGTCATATAAATGAATGAATA 110
Db 1 AAGGACATCTGCACCCAGTTTGGAACTGCAGTTTAAAGTCATATAAATGAATGAATA 60

QY 111 TGATAGCAAGGTGGAGGTGTTTAAAGAGCTATTTATAGGTCCTCGACAGCATCTTTT 170
Db 61 TGATAGCAAGGTGGAGGTGTTTAAAGAGCTATTTATAGGTCCTCGACAGCATCTTTT 120

QY 171 TCAATTAGGAGCAACCTTTTGGCCATGCGGTAACTGTGTGCAACTTCCTCTAAT 230
Db 121 TCAATTAGGAGCAACCTTTTGGCCATGCGGTAACTGTGTGCAACTTCCTCTAAT 180

QY 231 TGGGAATAGTTAAGCAGATCATAGAGCTGAATGATAAATTTGCTACGAGATGCACT 290
Db 181 TGGGAATAGTTAAGCAGATCATAGAGCTGAATGATAAATTTGCTACGAGATGCACT 240

QY 291 GGGACTCAACGTGACCTTATCAAGTGAGATGGAGTCTTGCCCTGCTCCAGGCTGGAGC 350
Db 1 GGGACTCAACGTGACCTTATCAAGTGAGATGGAGTCTTGCCCTGCTCCAGGCTGGAGC 1258

```

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu
----- Summary Statistics -----
Center project name: H_RG343H22

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

THE SEQUENCE OF THIS CLONE WAS ESTABLISHED AS PART OF A MAPPING AND SEQUENCING COLLABORATION BETWEEN THE NHGRI CHROMOSOME 7 MAPPING PROJECT AND THE WASHINGTON UNIVERSITY GENOME SEQUENCING CENTER. FOR ADDITIONAL INFORMATION ABOUT THE MAP POSITION OF THIS SEQUENCE, SEE <http://www.nhgri.nih.gov/DIR/GTB/CHR7> OR <mailto:egreen@nhgri.nih.gov>

SOURCE INFORMATION:

Clone CTA-343H22 is from a release of the human BAC library CITB-HS-A. The library contains cloned DNA from human sperm. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>).

VECTOR: pBelobAC11

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is RP11-573B16. Actual start of this clone is at base position 1 of CTA-343H22 actual end is at base position 131458 of CTA-343H22.

FEATURES

FEATURES	SOURCE
location_qualifiers	1. .131458
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="7"
	/map="7"
	/clone="CTA-343H22"
	/clone_lib="CITB-HS-A"
	32. .236
repeat_region	/rpt_family="ERV1"
repeat_region	256. .523
repeat_region	/rpt_family="MaLR"
repeat_region	1137. .1227
repeat_region	/rpt_family="ERV1"
repeat_region	1228. .1485
repeat_region	/rpt_family="Alu"
repeat_region	1486. .1897
repeat_region	/rpt_family="ERV1"
repeat_region	2406. .2489
repeat_region	/rpt_family="Tip100"
repeat_region	2721. .2877
repeat_region	/rpt_family="L1"
repeat_region	3370. .3614
repeat_region	/rpt_family="L1"
repeat_region	3619. .4195
repeat_region	/rpt_family="L1"
repeat_region	4730. .5239
repeat_region	/rpt_family="ERVU"
variation	4959
	/allele="T"

1431	Qy	TTCAACAAATATATCTACACATCAAAATTATAGATTGTTTTTCATTAATAATGCTTCATCT	1490
1259	Db	TTCAACAAATATATCTACACATCAAAATTATAGATTGTTTTTCATTAATAATGCTTCATCT	1318
1491	Qy	TAAACAGAATTGCTTTGTGTGATTGTTTTTAGAAACCTGAGAGTTTTTAATTCATTAATTACTT	1550
1319	Db	TAAACAGAATTGCTTTGTGTGATTGTTTTTAGAAACCTGAGAGTTTTTAATTCATTAATTACTT	1378
1551	Qy	GATCAAAAATTTGTGGGAACAATCCAGCATTAATTTGTAICTGATGTTGTTTTTATGTACATA	1610
1379	Db	GATCAAAAATTTGTGGGAACAATCCAGCATTAATTTGTAICTGATGTTGTTTTTATGTACATA	1438
1611	Qy	AGGAGTCTTAAAGCTTTGGTGCCTTGAAGCTTTTGTACTTAAGTCCCATGTTTAAAAATTACT	1670
1439	Db	AGGAGTCTTAAAGCTTTGGTGCCTTGAAGCTTTTGTACTTAAGTCCCATGTTTAAAAATTACT	1498
1671	Qy	ACTTTATATCTAAAGCATTTATGTTTTTCAAATTCAAATGATGCTAAATTAATGGCAA	1730
1499	Db	ACTTTATATCTAAAGCATTTATGTTTTTCAAATTCAAATGATGCTAAATTAATGGCAA	1558
1731	Qy	TTATACAATAATTAAGATTTCGAAATAGATA	1764
1559	Db	TTATACAATAATTAAGATTTCGAAATAGATA	1592

RESULT 6	AC002386	131458 bp	DNA	linear	PRI 29-APR-2003
LOCUS	AC002386	131458 bp	DNA	linear	PRI 29-APR-2003
DEFINITION	AC002386	131458 bp	DNA	linear	PRI 29-APR-2003
ACCESSION	AC002386	131458 bp	DNA	linear	PRI 29-APR-2003
VERSION	AC002386.2	GI:19034024	DNA	linear	PRI 29-APR-2003
KEYWORDS	HTG.	GI:19034024	DNA	linear	PRI 29-APR-2003
SOURCE	Homo sapiens (human)	GI:19034024	DNA	linear	PRI 29-APR-2003
ORGANISM	Homo sapiens	GI:19034024	DNA	linear	PRI 29-APR-2003
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	GI:19034024	DNA	linear	PRI 29-APR-2003
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	GI:19034024	DNA	linear	PRI 29-APR-2003
TITLE	1 (bases 1 to 131458)	GI:19034024	DNA	linear	PRI 29-APR-2003
JOURNAL	Sulston, J.E. and Waterston, R.	GI:19034024	DNA	linear	PRI 29-APR-2003
MEDLINE	Toward a complete human genome sequence	GI:19034024	DNA	linear	PRI 29-APR-2003
PUBMED	Genome Res. 8 (11), 1097-1108 (1998)	GI:19034024	DNA	linear	PRI 29-APR-2003
REFERENCE	99063792	GI:19034024	DNA	linear	PRI 29-APR-2003
AUTHORS	Waterston, R.	GI:19034024	DNA	linear	PRI 29-APR-2003
TITLE	2 (bases 1 to 131458)	GI:19034024	DNA	linear	PRI 29-APR-2003
JOURNAL	Cordes, M.	GI:19034024	DNA	linear	PRI 29-APR-2003
MEDLINE	The sequence of Homo sapiens BAC clone CTA-343H22	GI:19034024	DNA	linear	PRI 29-APR-2003
PUBMED	Unpublished (2001)	GI:19034024	DNA	linear	PRI 29-APR-2003
REFERENCE	3 (bases 1 to 131458)	GI:19034024	DNA	linear	PRI 29-APR-2003
AUTHORS	Waterston, R.	GI:19034024	DNA	linear	PRI 29-APR-2003
TITLE	Direct Submission	GI:19034024	DNA	linear	PRI 29-APR-2003
JOURNAL	Submitted (23-JUL-1997)	GI:19034024	DNA	linear	PRI 29-APR-2003
MEDLINE	4 (bases 1 to 131458)	GI:19034024	DNA	linear	PRI 29-APR-2003
PUBMED	Waterston, R.	GI:19034024	DNA	linear	PRI 29-APR-2003
REFERENCE	Direct Submission	GI:19034024	DNA	linear	PRI 29-APR-2003
AUTHORS	Submitted (03-FEB-2000)	GI:19034024	DNA	linear	PRI 29-APR-2003
TITLE	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA	GI:19034024	DNA	linear	PRI 29-APR-2003
JOURNAL	5 (bases 1 to 131458)	GI:19034024	DNA	linear	PRI 29-APR-2003
MEDLINE	Waterston, R.	GI:19034024	DNA	linear	PRI 29-APR-2003
PUBMED	Direct Submission	GI:19034024	DNA	linear	PRI 29-APR-2003
REFERENCE	Submitted (04-FEB-2000)	GI:19034024	DNA	linear	PRI 29-APR-2003
AUTHORS	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA	GI:19034024	DNA	linear	PRI 29-APR-2003
TITLE	6 (bases 1 to 131458)	GI:19034024	DNA	linear	PRI 29-APR-2003
JOURNAL	Waterston, R.	GI:19034024	DNA	linear	PRI 29-APR-2003
MEDLINE	Direct Submission	GI:19034024	DNA	linear	PRI 29-APR-2003
PUBMED	Submitted (01-MAR-2002)	GI:19034024	DNA	linear	PRI 29-APR-2003
REFERENCE	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA	GI:19034024	DNA	linear	PRI 29-APR-2003
AUTHORS	7 (bases 1 to 131458)	GI:19034024	DNA	linear	PRI 29-APR-2003
TITLE	Waterston, R.	GI:19034024	DNA	linear	PRI 29-APR-2003
JOURNAL	Direct Submission	GI:19034024	DNA	linear	PRI 29-APR-2003
MEDLINE	Submitted (01-MAR-2002)	GI:19034024	DNA	linear	PRI 29-APR-2003
PUBMED	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA	GI:19034024	DNA	linear	PRI 29-APR-2003
REFERENCE	8 (bases 1 to 131458)	GI:19034024	DNA	linear	PRI 29-APR-2003
AUTHORS	Waterston, R.	GI:19034024	DNA	linear	PRI 29-APR-2003
TITLE	Direct Submission	GI:19034024	DNA	linear	PRI 29-APR-2003
JOURNAL	Submitted (29-APR-2003)	GI:19034024	DNA	linear	PRI 29-APR-2003
MEDLINE	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA	GI:19034024	DNA	linear	PRI 29-APR-2003
PUBMED	On Mar 1, 2002 this sequence version replaced gi:2275181.	GI:19034024	DNA	linear	PRI 29-APR-2003
REFERENCE	----- Genome Center	GI:19034024	DNA	linear	PRI 29-APR-2003
AUTHORS		GI:19034024	DNA	linear	PRI 29-APR-2003
TITLE		GI:19034024	DNA	linear	PRI 29-APR-2003
JOURNAL		GI:19034024	DNA	linear	PRI 29-APR-2003
MEDLINE		GI:19034024	DNA	linear	PRI 29-APR-2003
PUBMED		GI:19034024	DNA	linear	PRI 29-APR-2003
REFERENCE		GI:19034024	DNA	linear	PRI 29-APR-2003
AUTHORS		GI:19034024	DNA	linear	PRI 29-APR-2003
TITLE		GI:19034024	DNA	linear	PRI 29-APR-2003
JOURNAL		GI:19034024	DNA	linear	PRI 29-APR-2003
MEDLINE					

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 193041)
Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Raymond,C. and
Haugen,E.D.
Direct Submission
Unpublished
2 (bases 1 to 193041)
Kaul,R.K., Olson,M.V., Raymond,C., Clendenning,J., Ivey,R.G. and
Haugen,E.D.
Direct Submission
Submitted (14-JUN-2001) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
3 (bases 1 to 193041)
Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Raymond,C. and
Haugen,E.D.
Direct Submission
Submitted (20-NOV-2001) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
On Nov 20, 2001 this sequence version replaced gi:14423605.
----- Genome Center
Center: University of Washington Genome Center
Center Code: UWGC
Web site: <http://www.genome.washington.edu>
Contact: uwgchgs@u.washington.edu
Drafting Center: SC
----- Project Information
Center project name: chr-1
Center clone name: RP11-543B16 (sc0173)
----- Summary Statistics
Sequencing vector: Plasmid; L08752; 100% of reads
Chemistry: Dye-terminator ET; 88% of reads
Chemistry: Dye-terminator Big Dye; 12% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 192648 bases at least Q40
Consensus quality: 133013 bases at least Q30
Consensus quality: 193041 bases at least Q20
Insert size: 200871; 9.7% error; agarose-fp
Insert size: 193041; sum-of-contigs
Quality coverage: 6.5x in Q20 bases; agarose-fp
Quality coverage: 6.7x in Q20 bases; sum-of-contigs

Overlapping Sequences:
5': Mapping in progress
3': Mapping in progress

Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
all regions were either double-stranded or sequenced with an
alternate chemistry or covered by high quality data (i.e., Phred
quality >= 30); an attempt was made to resolve all sequencing
problems, such as compressions and repeats; all regions were
covered by at least one plasmid subclone or more than one M13
subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:
This sequence has been validated by Multiple Complete Digest
fingerprinting. Comparison of the experimentally derived digest
fragments with sequence-predicted fragments is given below.
The electronically-digested sequence consists of both insert and
vector, in order to accurately represent the entire circular BAM.
Small fragments below a variable cutoff (approximately 400-800 bp)
are not resolved in the fingerprint and hence do not appear
in the table. There are no significant remaining discrepancies
between the experimental and predicted values. Unusually ordered

source

BASE COUNT	82426 a	54676 c	57516 g	101002 t	1615 others
ORIGIN					

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WormPEP; Information on the WormPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/ChrX>
RP13-310B24 is from the library RPCI-13.2 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>
VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP13-310B24. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
The true left end of clone RP13-310B24 is at 1 in this sequence.
The true left end of clone RP13-348B13 is at 99533 in this sequence.

FEATURES

	Location/Qualifiers	Source
	1..99632	
	/organism="Homo sapiens"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:9606"	
	/chromosome="X"	
	/map="q21.31-21.33"	
	/clone="RP13-310B24"	
	/clone_lib="RPCI-13.2"	
repeat_region	1..387	
	/note="TIGER1 repeat: matches 1175..1558 of consensus"	
repeat_region	388..692	
	/note="AluYa5 repeat: matches 1..305 of consensus"	
repeat_region	693..1810	
	/note="TIGER1 repeat: matches 45..1175 of consensus"	
repeat_region	1818..2361	
	/note="LIM2 repeat: matches 669..122 of consensus"	
repeat_region	3019..3339	
	/note="LIPB2 repeat: matches 5825..6154 of consensus"	
repeat_region	3395..5413	
	/note="LIM4 repeat: matches 4149..6270 of consensus"	
repeat_region	6911..7139	
	/note="L2 repeat: matches 1602..1829 of consensus"	
repeat_region	7232..7358	
	/note="MER91B repeat: matches 2..123 of consensus"	
repeat_region	7374..7711	
	/note="L2 repeat: matches 1826..2170 of consensus"	
repeat_region	8378..8873	
	/note="L2 repeat: matches 2176..2697 of consensus"	
repeat_region	9185..9476	
	/note="AluY repeat: matches 1..302 of consensus"	
repeat_region	9770..10131	
	/note="MER66A repeat: matches 2..433 of consensus"	
misc_feature	10137..10243	
	/note="match: STS: Em: L41959"	
repeat_region	11311..11561	
	/note="Tigger3(Golem) repeat: matches 2765..3027 of consensus"	
repeat_region	12387..12514	
	/note="LIMC1 repeat: matches 6175..6315 of consensus"	
repeat_region	12514..12691	
	/note="LIMC3 repeat: matches 7398..7586 of consensus"	

repeat_region	12703..12997	
	/note="AluY repeat: matches 1..292 of consensus"	
repeat_region	12999..13394	
	/note="MER31A repeat: matches 7..445 of consensus"	
repeat_region	13398..13903	
	/note="MT2CB repeat: matches 21..491 of consensus"	
repeat_region	13936..14701	
	/note="LIMC3 repeat: matches 6625..7403 of consensus"	
repeat_region	14702..14795	
	/note="47 copies 2 mer at 64% conserved"	
repeat_region	14799..14972	
	/note="LIMC3 repeat: matches 6407..6585 of consensus"	
repeat_region	14973..15384	
	/note="MSTA repeat: matches 1..426 of consensus"	
repeat_region	15385..17011	
	/note="MSTA-internal repeat: matches 1..1651 of consensus"	
repeat_region	17015..17356	
	/note="MSTA repeat: matches 4..359 of consensus"	
repeat_region	17357..17416	
	/note="30 copies 2 mer ta 88% conserved"	
repeat_region	17474..17531	
	/note="29 copies 2 mer ac 89% conserved"	
repeat_region	17534..17585	
	/note="MSTA repeat: matches 375..426 of consensus"	
repeat_region	17609..18436	
	/note="LIMC3 repeat: matches 5521..6388 of consensus"	
repeat_region	18438..19259	
	/note="LIP47 repeat: matches 5321..6141 of consensus"	
repeat_region	19283..19303	
	/note="LIMC3 repeat: matches 5379..5398 of consensus"	
repeat_region	19304..19687	
	/note="MLT1A1 repeat: matches 1..365 of consensus"	
repeat_region	19688..20527	
	/note="LIMC3 repeat: matches 4648..5379 of consensus"	
repeat_region	20528..20827	
	/note="AluX repeat: matches 1..300 of consensus"	
repeat_region	20828..21096	
	/note="LIMC3 repeat: matches 4390..4648 of consensus"	
repeat_region	21172..21750	
	/note="LIP16 repeat: matches 5532..6157 of consensus"	
repeat_region	21752..22295	
	/note="MLT2D repeat: matches 1..510 of consensus"	
repeat_region	22296..24134	
	/note="LIP45 repeat: matches 4291..6141 of consensus"	
repeat_region	24125..24386	
	/note="LIP45 repeat: matches 4017..4278 of consensus"	
repeat_region	24390..24457	
	/note="MLT2D repeat: matches 486..553 of consensus"	
repeat_region	24458..25607	
	/note="LIM1 repeat: matches 4410..5403 of consensus"	
repeat_region	25731..26237	
	/note="LIM4 repeat: matches 2749..3269 of consensus"	
repeat_region	26266..26491	
	/note="LIM4 repeat: matches 11..236 of consensus"	
repeat_region	26668..27305	
	/note="LIM1 repeat: matches 1390..750 of consensus"	
repeat_region	27306..27608	
	/note="AluSq repeat: matches 1..304 of consensus"	
repeat_region	27609..28895	
	/note="LIM1 repeat: matches 750..525 of consensus"	
repeat_region	28900..28988	
	/note="L1 repeat: matches 3816..3910 of consensus"	
repeat_region	28989..29293	
	/note="AluSq repeat: matches 1..305 of consensus"	
repeat_region	29294..29448	
	/note="L1 repeat: matches 3910..4043 of consensus"	
repeat_region	29450..29484	
	/note="MER61E repeat: matches 531..565 of consensus"	
repeat_region	29505..30294	
	/note="LIM3 repeat: matches 5176..5983 of consensus"	
repeat_region	32806..32901	
	/note="MIR repeat: matches 122..222 of consensus"	
repeat_region	33660..33985	

```

repeat_region /note="L2 repeat: matches 1737. .2088 of consensus"
34020. .34085
repeat_region /note="L2 repeat: matches 2630. .2700 of consensus"
34300. .34740
repeat_region /note="L2 repeat: matches 2281. .2710 of consensus"
34842. .35254
misc_feature /note="L1R16A repeat: matches 30. .445 of consensus"
35432. .35465
repeat_region /note="match: GSS: Em:AQ538246"
35466. .35840
repeat_region /note="MER39 repeat: matches 2. .385 of consensus"
complement(36265. .36420)
misc_feature /note="match: STS: Em:L77291"
38137. .38301
repeat_region /note="L1M10 repeat: matches 6157. .6318 of consensus"
38328. .39092
repeat_region /note="L1M81 repeat: matches 5396. .6158 of consensus"
39097. .39225
repeat_region /note="MER50 repeat: matches 14. .168 of consensus"
39320. .39739
repeat_region /note="MER50 repeat: matches 281. .712 of consensus"
39756. .40753
repeat_region /note="L1 repeat: matches 4445. .5419 of consensus"
40754. .41088
repeat_region /note="MER61A repeat: matches 2. .354 of consensus"
41090. .43719
repeat_region /note="MER61-internal repeat: matches 1959. .4550 of
consensus"
43720. .44403
repeat_region /note="L1R12 repeat: matches 1. .671 of consensus"
44406. .46188
repeat_region /note="MER61-internal repeat: matches 1. .1960 of
consensus"
46189. .46482
repeat_region /note="MER61A repeat: matches 13. .354 of consensus"
46490. .48685
repeat_region /note="L1 repeat: matches 2172. .4446 of consensus"

Query Match 9.5%; Score 494.6; DB 9; Length 99632;
Best Local Similarity 80.1%; Pred. No. 4.1e-79;
Matches 621; Conservative 0; Mismatches 144; Indels 10; Gaps 3;

QY 3323 TTTTAACTATGAAGTATGTTATCTGTATCTTTCATATAGTGTCCCGGAATTTATTT 3382
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 43678 TTTACAATAACGTAACATAAAACCTTAATATGATTGATAGTGTCCAGATTGGTTC 43737
QY 3383 CTTCCTGGTGGTCTTGTGTCGCTGACTCCAAAGTAAGAACCGCAGACCCCTTGAGGTGA 3442
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 43738 TTTCTGATGGTCTTGTGCTCTCACTGACTTCAAGAAATGAAGCGCAGACCTCACTGTGA 43797
QY 3443 GTGTACAGTCTTAAAGATGTTGTTTCAGAGTTGTTCTTCAGATGTTTCAGATGTTGT 3502
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 43798 GGTTCACAGTCTTAAAGATGTTGTTGTCAGTTGTTCTTCAGATGTTTCAGATGTTGT 43857
QY 3503 CCGGAGTTTCTCCCTTATGGTGAAGTTCGTGCTGCTGACTTCAACATGAAGCCGCGAG 3562
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 43858 CCAGAGTTTCTCCCTTCCAGTGGTTCATGCTGCTGCTGACTTCAAGAGTGAAGCCACAG 43917
QY 3563 ACCTTTGAGTGAAGTGTGACAGTCTTAAGAGCAGTGGTCCAGAGTGTGTTTCTCT 3622
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 43918 ACCTTTGAGTGA--GTGTTACAGCTTAAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 43975
QY 3623 CCGGTAGTGTGCTGCTGCTGATGTCAGGAATGAAGTGCAGACCCCTCGCGTAAGT 3682
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 43976 CTTGTGGTGGTGGTGAATCTTGCTGACTTTAGAGTGAAGCCACAGCCCTTCACG---GT 44031
QY 3683 GTTACAGCTCATAAAGTAGTGCACAAACCAACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 3742
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 44032 GTTACAGCTCATAAAGTAGTGCAGCCCAACCAAGGTTGAGCAGCAGATTTGTTATGA 44091
QY 3743 AGAGCAAAAGAACAAAGTCTCCCAACCATAGAACCGACCAAGATTGGTGTGCTGCTG 3802
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 44092 AGAGCAAAAGAACAAAGTCTCCACAGCATGGAAGGGGACTGAGCGGGTGTGCTGCTGCTG 44151

```

```

QY 3803 ----TGGTAGCCAGCTTTTATTCCCTTATTGGCCACACCCACATCCCTGCTGATGGCC 3858
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 44152 GCTCCGGTGGCCAGTTTATTATTCCTTATTGGCCACCCACACATCCCTGCTGATGGTCC 44211
QY 3859 ATTTACAGAATGCTGATGGTCCATTTATATACGCTGCTGATGGTGGCTTTTACAGAG 3918
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 44212 ATTTTACAGAGTGTGATGGTCCATTTTACAGAGTCTCTGATTTGGTGGCTTTTACAGAG 44271
QY 3919 TGCTGATGGTGCATTTTACATCCCTTAGCTAGACAGAGTGTGATGGTGGCTTTAT 3978
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 44272 TGCTCATTTGGTGTGTTTACATCCCTTAGCTAGACAGAGTGTGATGGTGGCTTTAT 44331
QY 3979 AATCCTTTAGCTAGACACAAAAGTCTTACAACTCCCAACCCACAGAGTGTGCTGG 4038
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 44332 AATCCTTTAGCTAGACACAAAAGTCTTCCAACTCCCAACCCACAGAGTGTGCTGG 44391
QY 4039 CTTCACCTCTCTTAAGAAATTTAGGTTCAAAAGTTCATAAGTGTCTAAACTA 4093
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 44392 CTTCACCTTTCAAAAGCATATACTCAGACATTAGAATTTAGAAATCCCATACAA 44446

```

```

RESULT 11
HS30G7/c
LOCUS
DEFINITION
Human DNA sequence from clone RP1-30G7 on chromosome Xp11.3-11.4,
complete sequence.
ACCESSION
AL034402
VERSION
AL034402.9 GI:4586155
KEYWORDS
HTG.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 123637)
AUTHORS
Whitehead,S.
TITLE
Direct Submission
JOURNAL
Submitted (04-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Apr 16, 1999 this sequence version replaced gi:4582299.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----

```

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/projects/C-elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chrx> RPI-30G7 is from the library RPI-1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

```
FEATURES
  source
    VECTOR: pcYPAC2.
    Location/Qualifiers
      1..123637
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="RZPD:RPCIF704G0730"
        /db_xref="taxon:9606"
        /chromosome="X"
        /map="p11.3-11.4"
        /clone="RP1-3067"
        /clone_lib="RPCI-1"
BASE COUNT 34340 a 25317 c 25127 g 38853 t
ORIGIN
  Query Match 9.38; Score 484.4; DB 9; Length 123637;
  Best Local Similarity 82.8%; Pred. No. 2.7e-77;
  Matches 591; Conservative 0; Mismatches 116; Indels 7; Gaps 3;
Qy 3341 TGTATCTGATCTTTTCATATAAGTGTCCCGGAATTTATTTCTCTGTTGGTCTTGG 3400
Db 96935 TGTATCTCTAAATCTTATCTGCTGATTCCTTTTGTGCCAGATTGGTAGTCTTGG 96876
Qy 3401 TCTCGTGATCTCAAGAAATGAACCGCAGACCCCTTGAGTGAGTGTCACAGTCTTAAAG 3460
Db 96875 TCTCGTGACTTCAAGAAATGAACCCATGAGCCCTCTTTGGAGAGTGTACAGTCTTAAAG 96816
Qy 3461 ATGTTGTGTCACAGTTTGTCTTTCAGATGTTTCAGATGTCGCGAGTGTCTCCCTAT 3520
Db 96815 ATGTTGTGTCGAGTTTGTCTTTCAGATGTTTCAGATGTCGCGAGTGTCTCCCTAT 96756
Qy 3521 GGTGAGTCTGTTGCTCGCTGACTTCAACAATGAAGCCGACACCTTTTGCAGTGAGTGTG 3580
Db 96755 GGTGGGTTTGTGCTCGCTGAGCTCAGGAGTGAAGCTGACAGACATTTGCGGTGA--GTG 96698
Qy 3581 TGACAGTCTTAAAGGACGTCGCTCCAGAGTTGTTTCTCCCGTAGTTCGTGGTC 3640
Db 96697 TTATAGCTTTAAAGTGATGCGCTCGGAGTTGTTTCTCTCTGTTGGTTCATGGTT 96638
Qy 3641 TCGTGTATGTCAGGAATGAAGTGCAGACCCCTCGCGTAAAGTGTACAGTCTATAAAGGT 3700
Db 96637 TCTGTGCTTCAGAGTGAAGTGCAGACCTTCAGGTGAGCGTTACAGTCTATAAAGG 96578
Qy 3701 AGTGCAAAACCAACAGTAGCAGTAGCAAGATTTATTATGAAGACAAAACAAGC 3760
Db 96577 AGCGCAGACCCAAACAGTAGCAGTAGCAAGATTTATTGTGAAGACCAAACAAGC 96518
Qy 3761 TTCCCAACCATAGAAACGGA--CCAGATTGGTTGCTGCTGCTG----TGTAGCCAGCTT 3815
Db 96517 TTCCACAGTACAGAAATGGGACCCAGACAGGTTGCTGCTGCTGCTGCTGCGCTGCTT 96458
Qy 3816 TTATTCCTTTATTTGGCCACACCCACATCTCTGCTGATTGGCCCATTTTACAGATGCTGA 3875
Db 96457 TTATTCCTTTATTTGGCCACACCCACATCTCTGCTGATTGGCCCATTTTACAGAGCTGA 96398
Qy 3876 TTGGTCCATTTTATAGCGTCTGATTTGGTCCGTTTTTACAGAGTGTGATTTGGTGCATTT 3935
Db 96397 TTTGTCCATTTTACAGAGTCTGATTTGGTCCGTTTTTACAGAGTGTGATTTGGTGCATTT 96338
Qy 3936 ACAATCTTTTAGTAGACAGAGTCTGATTTGGTCCGTTTTTATATCTTTTAGCTAGACACA 3995
Db 96337 ACAATCTTTTAGTAGACAGAGCGCTGATTTGGTGCATTTTAGCATCTTTTAGCTAGACA 96278
Qy 3996 CAAAAGTCTCAAGTCCCCACCAACCCAGAGAGTCCGCTGGCTTCCACCTCTC 4049
Db 96277 GAAAAGTCTCAAGTCCCCACCCAGAGAGTCCAGCCAGCTTCCACATCTC 96224
RESULT 12
AL137015 99527 bp DNA linear PRI 04-MAR-2003
LOCUS
DEFINITION Human DNA sequence from clone RP13-272N16 on chromosome
xq21.2-21.33, complete sequence.
ACCESSION AL137015
```


Qy 3881 CCATTTTATAGCGTGTGATTTGGTGGCTTTTACAGAGTGTGATTTGGTGCATTACAAT 3940
 Db 116886 CCATTTTACAGAGTTCGATTTGGTGGCTTTTACAGAGTGTGATTTGGTGCATTACAAT 116945
 Qy 3941 CCTTTAGCTAGACACAGAGTGTGATTTGGTGGCTTTTATATCCTTTAGCTAGACACAAA 4000
 Db 116946 ACTTTCCCTAGACACAGAGTGTGATTTGGTGCATTTTACATCTTTTACCTAGACACAAA 117005
 Qy 4001 GTTCTACAAGTCCCGCCACCCAGAGTCCGCTGGCTTCACCTCTC 4049
 Db 117006 GTTCTCAAGTCCCGCCACCCAGAGTCCGCTGGCTTCACCTCTC 117054

RESULT 14
 AC104697/c
 LOCUS
 DEFINITION Homo sapiens BAC clone RP11-454B3 from 2, complete sequence.
 AC104697 AC024403
 VERSION AC104697.2 GI:18855149
 KEYWORDS HTG.

SOURCE Homo sapiens (human)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 159797)
 TATCCTAGCTAGACACAGAGTGTGATTTGGTGGCTTTTATATCCTTTAGCTAGACACAAA
 Sulston, J.E. and Waterston, R.
 Toward a complete human genome sequence
 Genome Res. 8 (11), 1097-1108 (1998)
 95063792
 9847074

REFERENCE 2 (bases 1 to 159797)
 Radonenko, M. and Cotton, M.
 The sequence of Homo sapiens BAC clone RP11-454B3
 Unpublished (2001)
 3 (bases 1 to 159797)
 Waterston, R.H.
 Direct Submission
 Submitted (19-DEC-2001) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 4 (bases 1 to 159797)
 Waterston, R.
 Direct Submission
 Submitted (21-FEB-2002) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Feb 21, 2002 this sequence version replaced gi:17933873.

----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc
 Contact: sapiens@watson.wustl.edu
 ----- Summary Statistics
 Center project name: H_NH0454B03
 Drafting Center: WIBR

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this

sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenoe, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>
 VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-337B18, 2000 bp overlap;
 the clone sequenced to the right is RP11-241F6, 2000 bp overlap.
 Actual start of this clone is at base position 167881 of RP11-337B18; actual end is at base position 2890 of RP11-241F6.

Polymorphisms have been identified between AC092632, AC096560, and AC024403.

The sequence of AC024403 has been incorporated into AC104697.

FEATURES	Location/Qualifiers
source	1..159797
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
	/clone="RP11-454B3"
	/clone_lib="RPCI-11"
	340..410
	/rpt_family="(TA)n"
	492..514
	/rpt_family="AT-rich"
	1713..1931
	/rpt_family="MIR"
	2128..2272
	/rpt_family="MERL_type"
	2873..2998
	/rpt_family="L2"
	3102..3144
	/rpt_family="T-rich"
	3113..3425
	/rpt_family="Alu"
	4125..4273
	/rpt_family="MERL_type"
	5225..5808
	/rpt_family="L1"
	5811..7833
	/rpt_family="L1"
	7088..7094
	/rpt_family="AT-rich"
	8088..8382
	/rpt_family="Alu"
	8742..8807
	/rpt_family="AT-rich"
	8900..9206
	/rpt_family="Alu"
	9180..9208
	/rpt_family="AT-rich"
	9230..9332
	/rpt_family="MIR"
	11500..11687
	/rpt_family="Alu"
	11680..11714
	/rpt_family="(TTTA)n"
	11688..11998
	/rpt_family="Alu"
	11999..12083
	/rpt_family="Alu"
	12622..13188
	/rpt_family="L1"

